



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Interrogation of mammalian protein complex structure, function, and membership using genome-scale fitness screens

**Citation for published version:**

Pan, J, Meyers, R, Michel, B, Mashtalir, N, Sizemore, AE, Wells, J, Cassel, SH, Vazquez, F, Weir, BA, Hahn, WC, Marsh, J, Tsherniak, A & Kadoch, C 2018, 'Interrogation of mammalian protein complex structure, function, and membership using genome-scale fitness screens', *Cell Systems*, vol. 6, no. 5, pp. 555-568.e7. <https://doi.org/10.1016/j.cels.2018.04.011>

**Digital Object Identifier (DOI):**

[10.1016/j.cels.2018.04.011](https://doi.org/10.1016/j.cels.2018.04.011)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Cell Systems

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

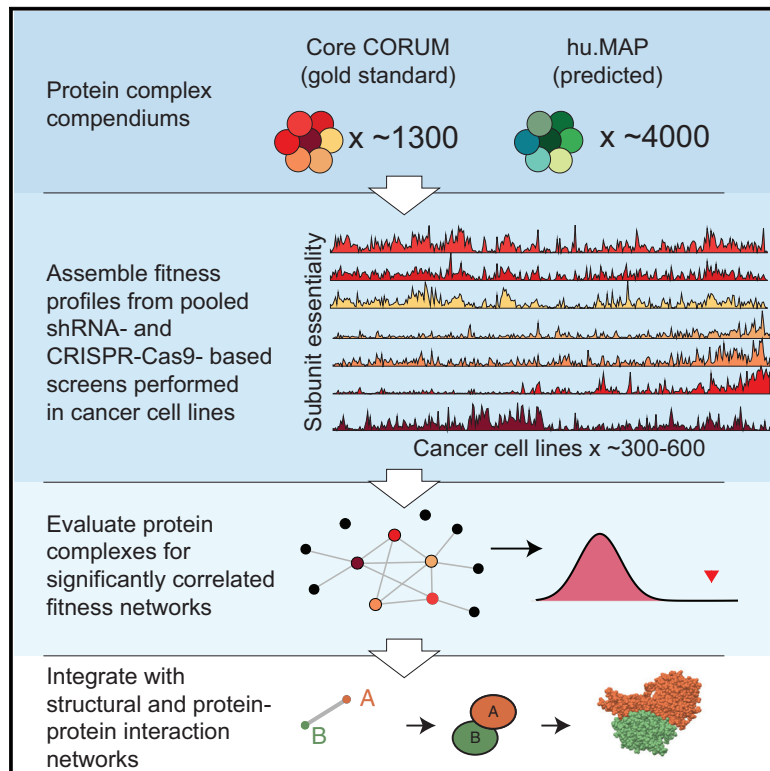
**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Interrogation of Mammalian Protein Complex Structure, Function, and Membership Using Genome-Scale Fitness Screens

## Graphical Abstract



## Authors

Joshua Pan, Robin M. Meyers, Brittany C. Michel, ..., Joseph A. Marsh, Aviad Tsherniak, Cigall Kadoch

## Correspondence

cigall\_kadoch@dfci.harvard.edu

## In Brief

Fitness profiling from large-scale genetic screens across cancer cell lines enables the analysis of protein complexes as genetic networks. Clustering subunits based on correlated fitness profiles reveals biochemical and functional modularity of protein complexes and allows for discovery of new complexes when integrated with protein-protein interaction predictions.

## Highlights

- Genome-scale fitness screens define correlation networks for protein complexes
- ~40% of Core CORUM complexes exhibit significantly correlated fitness profiles
- Fitness correlations resolve structural modules of multimeric protein complexes
- Integrated fitness and predicted PPI data nominate new functional interactions



# Interrogation of Mammalian Protein Complex Structure, Function, and Membership Using Genome-Scale Fitness Screens

Joshua Pan,<sup>1,2,3,11</sup> Robin M. Meyers,<sup>2,8,11</sup> Brittany C. Michel,<sup>1,2,3</sup> Nazar Mashtalir,<sup>1,2</sup> Ann E. Sizemore,<sup>2,9</sup> Jonathan N. Wells,<sup>4</sup> Seth H. Cassel,<sup>1,2,3,5</sup> Francisca Vazquez,<sup>2</sup> Barbara A. Weir,<sup>2,10</sup> William C. Hahn,<sup>1,2,6,7</sup> Joseph A. Marsh,<sup>4</sup> Aviad Tsherniak,<sup>2</sup> and Cigall Kadoch<sup>1,2,6,12,\*</sup>

<sup>1</sup>Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215, USA

<sup>2</sup>Broad Institute of Harvard and MIT, 415 Main Street, Cambridge, MA 02142, USA

<sup>3</sup>Biomedical and Biological Sciences Program, Harvard Medical School, Boston, MA 02115, USA

<sup>4</sup>MRC Human Genetics Unit, Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK

<sup>5</sup>Medical Scientist Training Program, Harvard Medical School, Boston, MA 02115, USA

<sup>6</sup>Harvard Medical School, Boston, MA 02115, USA

<sup>7</sup>Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

<sup>8</sup>Present address: Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA

<sup>9</sup>Present address: Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104, USA

<sup>10</sup>Present address: Janssen Research and Development, Cambridge, MA 02142, USA

<sup>11</sup>These authors contributed equally

<sup>12</sup>Lead Contact

\*Correspondence: [cigall\\_kadoch@dfci.harvard.edu](mailto:cigall_kadoch@dfci.harvard.edu)

<https://doi.org/10.1016/j.cels.2018.04.011>

## SUMMARY

Protein complexes are assemblies of subunits that have co-evolved to execute one or many coordinated functions in the cellular environment. Functional annotation of mammalian protein complexes is critical to understanding biological processes, as well as disease mechanisms. Here, we used genetic co-essentiality derived from genome-scale RNAi- and CRISPR-Cas9-based fitness screens performed across hundreds of human cancer cell lines to assign measures of functional similarity. From these measures, we systematically built and characterized functional similarity networks that recapitulate known structural and functional features of well-studied protein complexes and resolve novel functional modules within complexes lacking structural resolution, such as the mammalian SWI/SNF complex. Finally, by integrating functional networks with large protein-protein interaction networks, we discovered novel protein complexes involving recently evolved genes of unknown function. Taken together, these findings demonstrate the utility of genetic perturbation screens alone, and in combination with large-scale biophysical data, to enhance our understanding of mammalian protein complexes in normal and disease states.

## INTRODUCTION

The derivation of gene-gene relationships is a central goal of systems genetics (Baliga et al., 2017). Several gene properties

including co-temporal expression, co-evolution, and physical interaction between protein products have proven to be informative in the identification of functionally related genes such as those coding for subunits of protein complexes (de Juan et al., 2013; Gingras et al., 2007; Jansen et al., 2002; Ramani et al., 2008). One particularly powerful approach is to define genetic interactions by probing for epistatic relationships between genes for which the phenotypic readout of a genetic perturbation depends on the status of a second gene (Baryshnikova et al., 2013).

Genetic interaction mapping has been most extensively pursued in *S. cerevisiae*, in which crosses between gene knockout strains coupled with cellular fitness readouts enabled systematic measurements of genetic interactions (Pan et al., 2004; Schuldiner et al., 2005; Tong et al., 2004). Studies have indicated that genes functioning within similar biological processes tend to share genetic interaction partners (Collins et al., 2007; Kelley and Ideker, 2005; Schuldiner et al., 2005), which motivated the construction of genome-scale functional similarity networks for yeast (Costanzo et al., 2010; Costanzo et al., 2016). In these networks, functionally related genes share an edge based on the similarity of their genetic interaction profiles, ultimately yielding a modular, hierarchical model of the cell, in which genes with coordinated functions, such as members of the same protein complex, cluster into functional modules.

Given the utility of this approach for inferring gene function, a major interest in the field lies in deriving global functional similarity maps for human cells. However, the generation of such datasets faces major limitations. In addition to the roughly 16-fold increase in combinatorial space of all possible double knockouts in human cells compared with yeast (due to the 4-fold increase in genes), human screening libraries for simultaneously knocking out multiple genes of interest are relatively new and still face technical challenges (Boettcher et al., 2018; Du et al., 2017;



Han et al., 2017; Najm et al., 2018; Shen et al., 2017). In contrast, genome-scale single-gene knockout libraries are technically more advanced and have been used extensively in pooled fitness screens (Doench, 2018). In theory, performing an equivalent double-knockout experiment with a single-gene knockout library would require screening for fitness effects across a massive cell line collection, in which each cell line contains a precise genetic knockout stably derived from an isogenic background. Instead, we hypothesized that an analogous approach may be feasible using a collection of cancer cell lines with sufficient diversity in fitness responses, in which diversity does not arise from single knockouts but rather through genomic and transcriptomic variation en masse.

We and others have observed that cancer cell lines exhibit highly variable genetic dependencies for cellular fitness (Bertomeu et al., 2018; Blomen et al., 2015; Hart et al., 2015, 2017a; McDonald et al., 2017; Meyers et al., 2017; Tsherniak et al., 2017; Wang et al., 2015, 2017b) in a manner that reflects the diverse genomic and transcriptomic alterations a cell may accumulate during tumorigenesis. Project Achilles (Broad Institute) seeks to systematically map genetic vulnerabilities across large collections of cancer cell lines, including the Cancer Cell Line Encyclopedia (Barretina et al., 2012), and has recently performed genome-scale perturbation screens in 501 cancer cell lines using RNAi (Tsherniak et al., 2017) and in 342 cancer cell lines using CRISPR-Cas9 (Meyers et al., 2017). Because the genomic and transcriptomic state of each cancer cell line gives rise to a unique overall fitness response upon perturbation of each gene, these datasets may provide an opportunity to derive gene-gene functional relationships and to construct a modular network of cell function.

In this study, we evaluate the use of large-scale RNAi and CRISPR-Cas9 genetic perturbation datasets for the construction of a human functional similarity network. We focused on protein complexes because of their modular composition, coordinated function, and involvement in many biological processes (Pereira-Leal et al., 2006). Following validation that correlated fitness profiles between gene pairs in both RNAi- and CRISPR-Cas9-based screens represent informative measures of known functional relationships, such as physical interactions, we then developed a network permutation approach to evaluate whether gene modules coding for protein complex subunits are significantly correlated in their fitness profiles. After benchmarking against a gold standard protein complex dataset (Core CORUM), we find that ~40% of gold standard protein complexes form significantly connected functional modules in our networks. We find that these functional similarity networks reproduced structural features of known protein complexes, as well as identified intra-complex functional modularity in complexes with no known structure. Finally, we applied this approach to a set of computationally predicted but unvalidated protein complexes (hu.MAP), and describe a pair of functionally related but uncharacterized genes whose protein products form a novel protein complex. Taken together, these findings establish the utility of large-scale fitness screening in cancer cell lines to reveal functional and structural features of protein complexes, and prime the field for the global derivation of a human functional similarity network from large-scale genetic perturbation screens.

## RESULTS

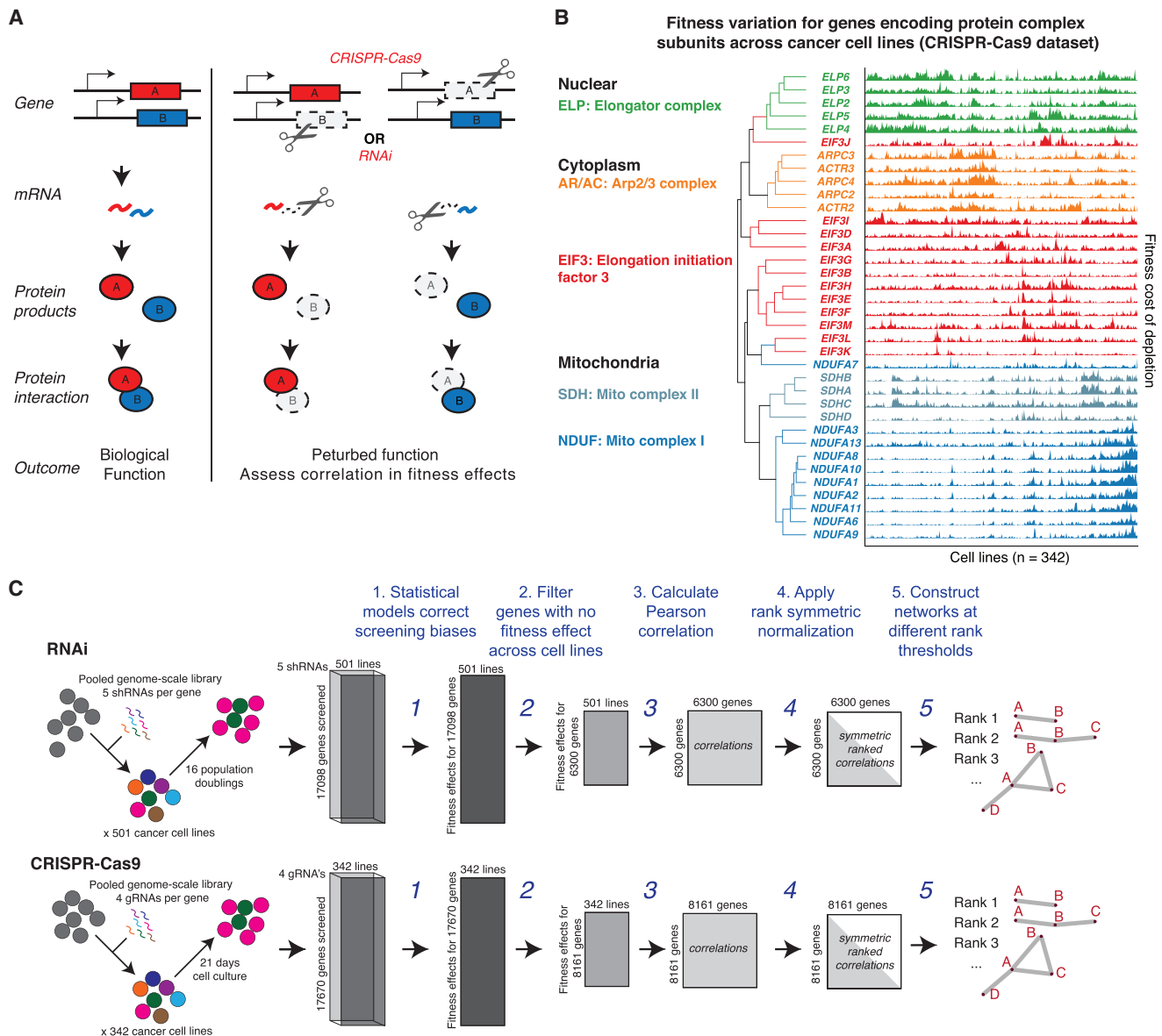
### Interacting Proteins Exhibit Coordinated Fitness Effects upon Genetic Depletion in Cancer Cell Lines

Protein complexes execute specific molecular functions that require the proper assembly and activity of their interacting subunits (Figure 1A, left). Depletion of individual subunits required for complex function would be predicted to produce similar phenotypic effects on fitness (Figure 1A, right). To assess this premise systematically in human cells, we analyzed recently generated datasets from large-scale RNAi- and CRISPR-Cas9-based fitness screening efforts of hundreds of cancer cell lines via Project Achilles (<https://portals.broadinstitute.org/achilles>) (Cowley et al., 2014; Meyers et al., 2017; Tsherniak et al., 2017). Over 600 cancer cell lines from 23 different lineages were screened across the RNAi and CRISPR-Cas9 datasets, representing an extremely diverse set of cellular contexts (Figure S1A and Table S1).

As an illustration of this premise, the fitness effects upon CRISPR-Cas9 knockout of subunits of various protein complexes across 342 cell lines are shown (Figure 1B). The fitness effects upon gene knock out varied widely across cell lines, and analysis of both RNAi and CRISPR-Cas9 datasets demonstrated that genes coding for Core CORUM protein complex subunits exhibited significantly greater fitness variation than genes that do not (Wilcoxon rank-sum test,  $p < 1e-3$ ) (Figure S1B; Table S2). In addition, fitness profiles for genes encoding subunits of the same protein complexes are strikingly concordant, and correlation clustering by fitness profile resulting in genes being grouped by protein complex membership (Figure 1B). Across both datasets, pairs of genes sharing Core CORUM protein complex membership exhibit significantly greater correlations of fitness profiles than protein subunits from different complexes (Figure S1C, Kolmogorov-Smirnov test,  $p = 2.2e-16$  for both RNAi and CRISPR-Cas9). This result suggests the potential to use correlated fitness effects across cancer cell lines as a measure of functional similarity, and therefore identify and resolve functional relationships within and between protein complexes.

To test this observation for interacting proteins in general, we collated large-scale, annotated protein-protein interaction (PPI) datasets generated by individual or joint research groups over the past 5 years (Drew et al., 2017; Hein et al., 2015; Huttlin et al., 2017; Rolland et al., 2014; Thul et al., 2017; Wan et al., 2015), including Core CORUM as the gold standard literature reference (Ruepp et al., 2010). We again found that gene pairs with an annotated PPI consistently exhibited greater fitness profile correlations across all datasets than gene pairs that did not have an annotated interaction (Figure S1D). Furthermore, after binning gene pairs by the strength of their fitness correlations in the RNAi and CRISPR datasets (as performed in Hart et al., 2017a, using the KEGG Pathway Database), we found that the most strongly correlated gene pairs were the most enriched for interacting protein products across PPI datasets (Figure S1E).

Taken together, these results demonstrate that interacting proteins tend to exhibit correlated fitness profiles in large-scale genetic perturbation screens, and that top-ranked fitness correlations were highly enriched for interacting proteins, suggesting that many protein complexes may exist as correlated modules



**Figure 1. Genes Encoding Protein Complex Subunits Display Coordinated Fitness Variation across Genetic Screens Performed in Human Cancer Cell Lines**

(A) Schematic of normal and perturbed protein complex biogenesis.

(B) Fitness profiles for genes encoding subunits of five different protein complexes screened in the CRISPR-Cas9 fitness dataset, annotated by their gene name abbreviations and cellular localization. Both rows (genes) and columns (cell lines) are hierarchically clustered.

(C) Graphical representation of RNAi- and CRISPR-Cas9-based screening datasets and analysis pipelines ( $n = 501$  and  $n = 342$  cell lines, respectively; Project Achilles, Broad Institute).

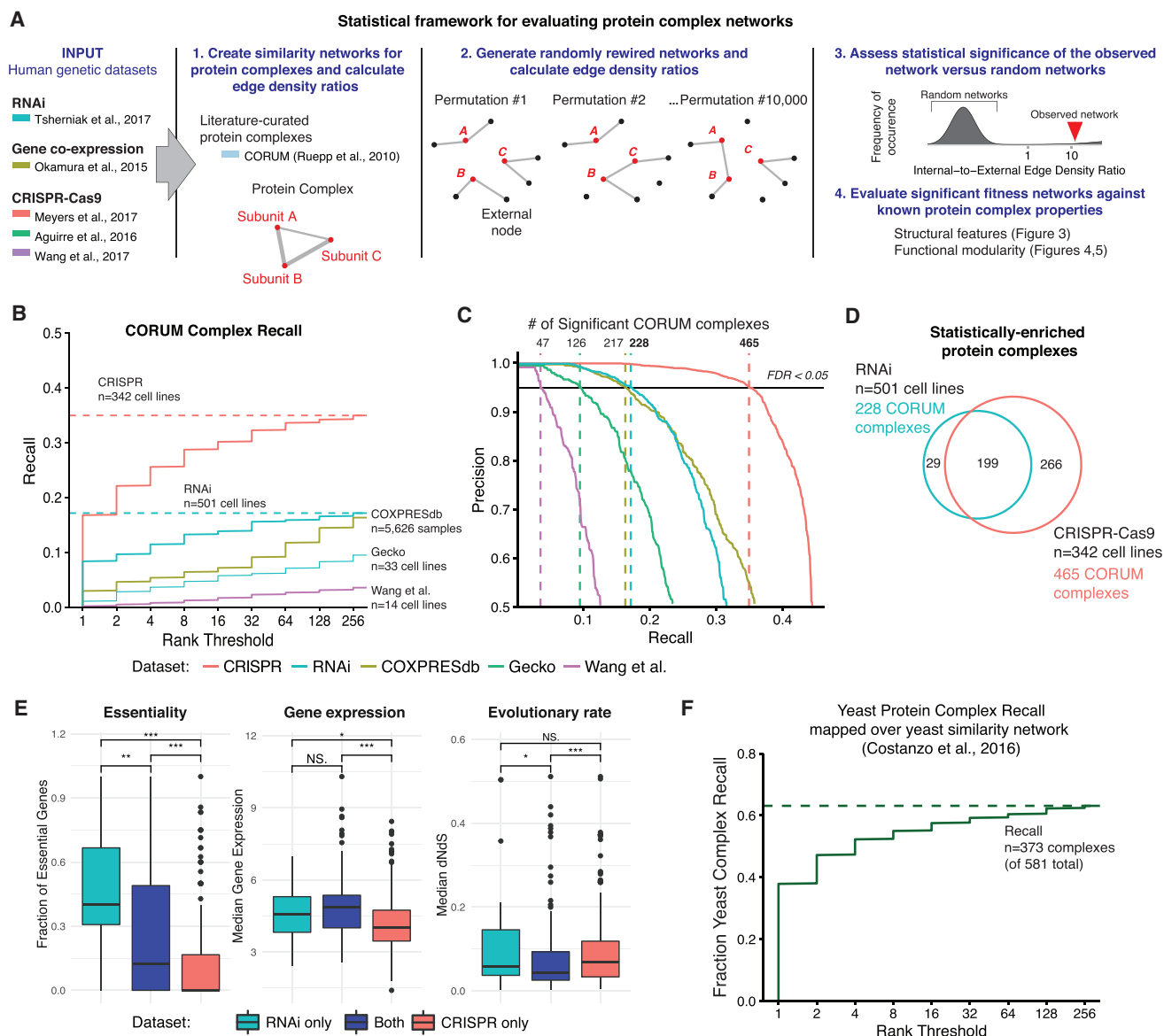
within these fitness screening datasets. To further test this, we analyzed both the RNAi and CRISPR-Cas9 screening data in parallel. From these screening data, all collected at one screening facility and subjected to statistical models that eliminate off target effects, we filtered for genes whose depletion had significant fitness effects to obtain 6,300 genes in the RNAi dataset and 8,161 genes in the CRISPR-Cas9 dataset (Figure 1C). We then performed Pearson correlations on fitness profiles for all pairwise combination of genes, rank-order normalized the vector of correlations for each gene, symmetrized by taking

the best rank across gene pairs, and used these data as the basis for all networks generated below.

### Using RNAi and CRISPR-Cas9 Screening Data to Generate Functional Similarity Networks for Hundreds of Human Protein Complexes

To systematically assess the extent to which fitness screening data can inform the biology of protein complexes, we developed a network-based approach for analyzing both RNAi and CRISPR fitness datasets (Figure 2A [summary], Figure S2 [detail], see the





**Figure 2. A Statistical Framework for Nominating Significant Protein Complex Fitness Correlation Networks**

(A) Overview of the statistical framework for identifying significant protein complex fitness correlation networks (see Figure S2).  
(B) Fraction of human protein complexes recalled at FDR < 0.05 in fitness correlation datasets (RNAi, CRISPR, Gecko, and Wang et al.) and a gene expression correlation dataset (COXPRESdb), plotted against a log range of rank correlation thresholds. Fraction of CORUM complex recall is defined as the fraction of CORUM protein complexes (n = 1,331) that exhibit correlations at or above that rank threshold.  
(C) Precision-recall curve for the protein complexes in each dataset.  
(D) Venn diagram depicting overlap between CORUM protein complexes statistically enriched with top-ranked correlations in CRISPR and RNAi datasets.  
(E) Biologic properties of protein complexes with significant correlations in RNAi, CRISPR, or both datasets. Lower and upper hinges of boxplots show first and third quartiles, respectively. Lower and upper whiskers extend to the smallest and largest values, respectively, that are within 1.5× the interquartile range (difference between first and third quartiles) from the nearest hinge. Wilcoxon rank-sum test, \*\*p < 1e-2, \*\*\*p < 1e-3, N.S., not significant.  
(F) Statistical framework in (A) applied to a yeast correlation dataset derived from a genome-scale pairwise interaction map (Costanzo et al., 2016). A cumulative total of 373 yeast protein complexes with statistically significant fitness networks were recalled at rank 256, representing 64% of total yeast protein complexes.

**STAR Methods**). We represent a protein complex as a functional similarity network consisting of nodes for each protein subunit and edges when the measure of functional similarity between two subunits exceeds a given threshold. We define functional similarity using the symmetric rank-normalized correlations as above. At any given rank threshold, edges may connect two sub-

units within a protein complex of interest (internal edges), or they may connect subunits between two different complexes (external edges). To determine the set of protein complexes whose subgraphs are statistically enriched for internal edges in this network, we calculated the ratio of internal edge density versus external edge density for each protein complex across

several rank thresholds (Figures S2A–S2C). We then determined the statistical significance of this ratio using an empirical null distribution generated from 10,000 randomly rewired networks while preserving node degree. Finally, we visualized the functional similarity network for protein complexes that exceed statistical significance (false discovery rate [FDR] < 0.05) (Figure S2D) and analyzed their functional and structural features.

To compare the performance of our methodology on the RNAi and CRISPR fitness datasets described above, we included two additional recently published CRISPR-Cas9 screening datasets: one consisting of 14 CRISPR-Cas9-based screens performed in acute myeloid lymphoma cancer lines (Wang et al., 2017b), and the other consisting of 33 cancer cell lines of diverse lineage screened with the GeCKOv2 sgRNA library (Aguirre et al., 2016) (Figures 2A, S3A, and S3B). In addition, to draw comparisons with previous studies of correlated gene expression profiles of protein complexes, we included a large-scale mRNA co-expression dataset, COXPRESdb (Okamura et al., 2015), containing pairwise correlations across 5,000+ publically available RNA sequencing datasets (Figure 2A). We generated the gene networks for each of these five datasets and determined the set of Core CORUM protein complexes whose subgraphs are significantly enriched for internal edges at various rank thresholds.

The CRISPR-Cas9 and RNAi fitness datasets both captured a greater fraction of human protein complexes than the smaller GeCKOv2 or Wang et al. datasets—likely due to the scale of cancer cell lines screened—and also outperformed the COXPRESdb network across top rank thresholds (Figures 2B and 2C). Specifically, by rank 256, 17% of all Core CORUM protein complexes were captured at statistical significance in the RNAi dataset and 35% of all complexes in the CRISPR dataset. The distribution of the cellular localization of protein complexes enriched in both RNAi or CRISPR datasets was similar to the localization distribution for all human protein complexes (Core CORUM), indicating that protein complexes from a variety of processes and pathways exhibit significant functional similarity networks (Figure S3C).

In total, we found that 494 out of 1,331 Core CORUM complexes had statistically significant fitness networks in either RNAi or CRISPR dataset, with 228 complexes from RNAi, 465 from CRISPR, and 199 overlapping between the two (Figure 2D; Table S3). We identified features of protein complexes that underpin differences in recall between the two genetic perturbation datasets. Protein complexes recalled only in the CRISPR dataset were significantly depleted of core essential genes (Hart et al., 2015, 2017b), had lower median gene expression levels, and had lower sequence conservation (Wilcoxon rank-sum test,  $p < 1e-3$ ) than those recalled in both RNAi and CRISPR datasets (Figure 2E; Table S3).

To further benchmark this performance against a large-scale functional mapping effort in a model organism, we generated the functional similarity network from a dataset generated by Costanzo et al. (2016) in their effort to map all pairwise genetic interactions in *S. cerevisiae*. Applying our methodology to this functional similarity network, which is derived by calculating pairwise correlations of genetic interaction profiles, we captured 64% (373/581) of all yeast complexes as significantly correlated subgraphs (Figure 2F).

We visualized individual functional similarity networks for each of the 494 Core CORUM protein complexes recalled in either dataset, first by determining the dataset in which that protein complex exhibits the greatest enrichment for internal edges (Figure S3D), and second by choosing a rank threshold specifically for the complex that optimizes the ratio of internal to external edges (see the STAR Methods). We found that these functional similarity networks vary greatly in their network topology (Figure S3E) and included many well-studied protein complexes (26S Proteasome, Mediator, RNA polymerase II, STAGA, mammalian SWI/SNF, and others) with structural or functional features that we sought to examine in greater depth.

### Functional Similarity Networks Recapitulate Structural and Functional Modules of Protein Complexes

Large protein complexes are often hierarchically assembled from smaller sub-assemblies that have specific modular functions. We hypothesized that for protein complexes with a sufficient number of subunits and sufficiently distinct functional componentry, this modular structure may be reflected in the functional similarity networks. The Mediator complex is an evolutionarily conserved transcriptional activator composed of three stable modules—the Head, Middle, and Tail (Figure 3A)—and one detachable, cell-cycle-specific module (cyclin kinase module, not crystallized) (Nozawa et al., 2017; Tsai et al., 2017). Recent biochemical and genetic studies have demonstrated that the Head, Middle and Tail modules exhibit differential genomic targeting and have specialized roles in the context of Mediator complex global function (Jeronimo et al., 2016; Petrenko et al., 2016). Depletion of subunits belonging to the Head, Middle and Tail modules results in distinct cell fitness effects across cancer cell lines (Figure 3B). Correspondingly, the CRISPR functional similarity network for the Mediator complex largely contains edges between subunits belonging to the same structural module at various rank thresholds (Figure 3C). Comparatively, a gene expression similarity network for the Mediator complex derived from the COXPRESdb dataset captured few correlations between the Head or Middle modules, even at a lenient rank threshold of 50 (Figure S4A).

We sought to understand whether or not there were additional protein complexes that displayed overlapping physical and functional modularity. After restricting our analysis to large protein complexes (ten or more subunits) that displayed significant fitness networks, we compared fitness network communities with those inferred from PPI networks, using the Bioplex 2.0 dataset and a PDB structural interaction network (Figures S4B and S4C; Table S4). In comparison with the 29 protein complexes displaying overlapping modularity between these structural interaction networks and hu.MAP (another protein interaction network), we found that 12 protein complexes showed significantly overlapping fitness network modules versus their structural counterparts, and only two displayed overlapping structural and mRNA co-abundance network modularity (Figure S4D and Table S5). Several of these protein complex assemblies, including the 26S proteasome, RNA Pol holoenzyme, and the COP9 signalosome (Figure S4D, inset), showed overlapping functional and physical modularity over recognizable sub-complexes and assembly modules.

The 26S proteasome is composed of the 19S regulatory and 20S core sub-assemblies joined by a common interface





(Figure 3D). The RNAi functional similarity network clearly distinguished the subunits of the 19S and 20S particles; of the 31 genes encoding proteasome subunits included in this dataset, 20 were involved in a top-ranked correlation with another proteasome subunit, largely within the same particle (Figure 3E, left). This is consistent with the fact that depletion of the 19S and 20S particles was previously shown to have differing effects on cancer cells, despite both particles being part of the same macromolecular assembly (Dambacher et al., 2016). While co-expression networks of 26S proteasome subunits are strongly connected, they fail to reach significance using a module overlap test for structural networks (Figure S4E).

Heteromeric complexes follow energetically favorable ordered assembly pathways, forming intermediate configurations during this process (Ahnert et al., 2015). The RNA Pol II subunits POLR2C and POLR2J form a heterodimer that acts as an assembly platform to nucleate the remainder of the assembly pathway, which occurs via three sub-assembly intermediates (Figure 3F) (Wild and Cramer, 2012), while the POLR2D-G detachable heterodimeric recognition stalk selectively associates with the complex via POLR2A during transcription elongation (Werner and Grohmann, 2011). The RNAi functional similarity network of the RNA Pol II complex identified the POLR2C-J and POLR2D-G heterodimers as rank 1 pairwise correlations (Figure 3G, left). The network further highlighted functional relationships between the sub-assembly three components that anchor the recognition stalk to the complex (Tan et al., 2003) and the assembly platform component POLR2K and its binding partner POLR2B, which assemble together on the POLR2C-J heterodimer (Figure 3G, right) (Wild and Cramer, 2012). The observed overlap between edges in the functional similarity network and the structural interaction network is statistically significant (two-sided Fisher's exact test,  $p = 8.9e-3$ ). The co-expression network for RNA Pol II does not resolve this functional modularity (Figure S4F).

To test the extent to which such heterodimers exhibit correlation in fitness profiles globally, we curated a set of 271 heterodimeric interactions from the full PDB structural interaction dataset (Table S4) and found significantly higher ranked correlations in both the RNAi and CRISPR datasets than in COXPRESdb dataset (Kolmogorov-Smirnov test, RNAi versus COXPRESdb,  $p = 4.7e-3$ , CRISPR versus COXPRESdb,  $p = 5.2e-4$ ; Figure S4G). Taken together, these results suggest that stably interacting, functionally obligate subunits, such as those in heterodimers or structural submodules of protein complexes, exhibit highly ranked fitness correlations in these fitness datasets.

### Functional Similarity Networks Resolve Shared Subunits with Functionally Diverged Protein Complexes

Novel complexes can form over evolutionary time via partial duplication and divergence of specific subunits, which can adopt new function while maintaining stable interactions with shared subunits from the original complex (Pereira-Leal et al., 2006). The eukaryotic RNA Pol I, II, and III arose according to this paradigm, with select subunits duplicating and diverging between Archaea and Eukaryotes (Carter and Drouin, 2009) (Figure S5A). Hierarchical clustering performed on fitness profiles of RNA Pol I, II, and III subunits in the RNAi dataset functionally distinguish RNA Pol III from RNA Pol I and II (Figure S5B). Interestingly, components of the heterodimeric POLR1C-D assembly

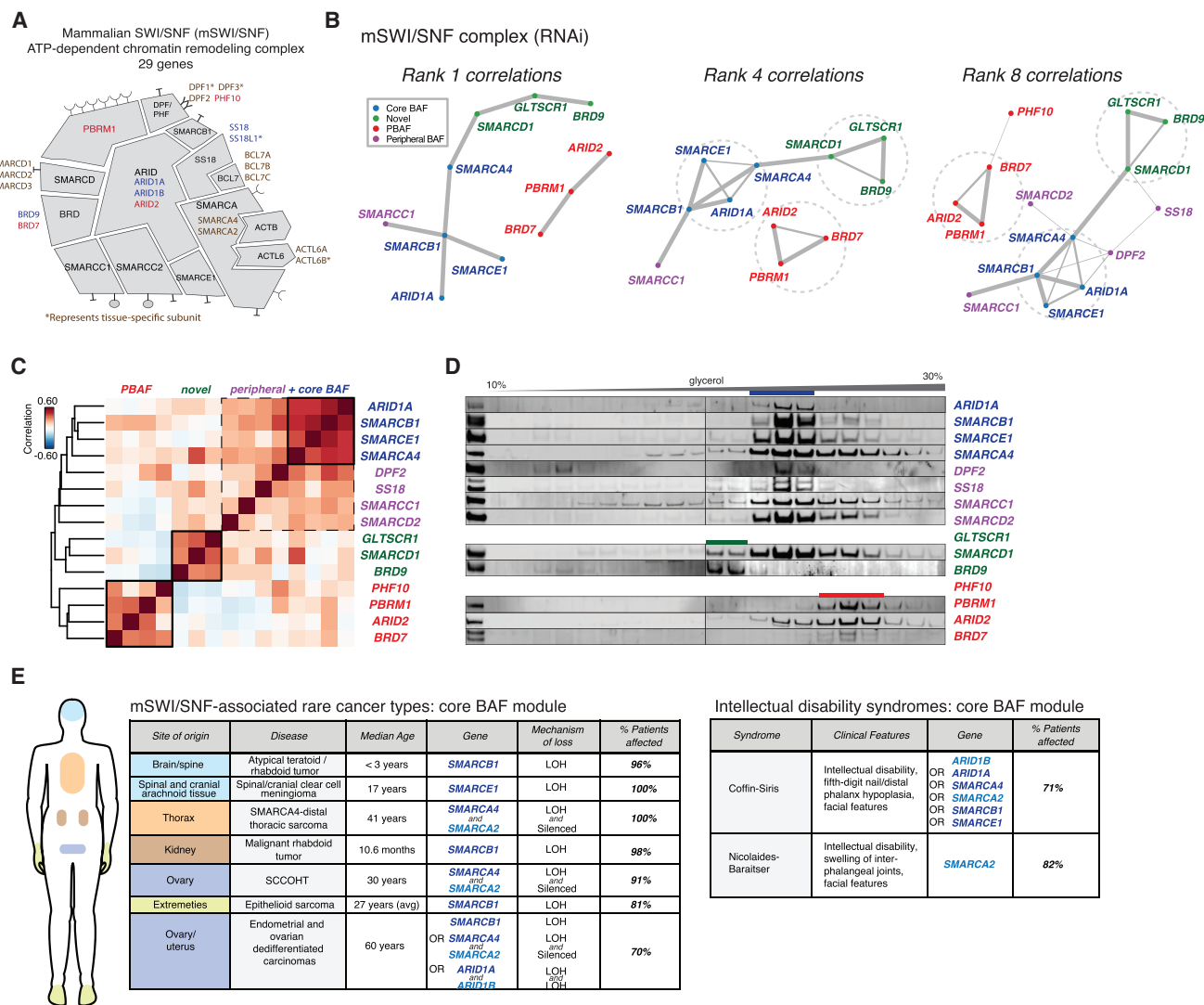
platform (homologous to POLR2C-J) show functional similarity to both Pol I and Pol III in the functional similarity network for the three polymerases (Figure S5C). This heterodimer is shared between both complexes, and blocking their dimerization precludes assembly of either complex in *S. cerevisiae* (Mann et al., 1987).

The functional similarity network of the STAGA/ATAC family of complexes, which are known to deposit acetylation marks genome wide (Spedale et al., 2012), reveals structural modularity and suggests distinct functional characteristics (Figure S5D). Although members of the STAGA complexes cluster distinctly from members of the ATAC complex in the CRISPR dataset (Figure S5E), three subunits shared between both complexes—KAT2A, CCDC101, and TADA3—appear centrally in the functional similarity network, bridging the two complexes (Figure S5F). The gene co-expression network did not recapitulate this modularity or shared subunit membership for the RNA Pol and the STAGA/ATAC family of complexes (Figures S5G and S5H). These results collectively demonstrate the resolution with which RNAi- and CRISPR fitness networks report on the modularity and assembly of protein complexes within defined complex families.

### Functional Similarity Identifies a Novel Functional Module of the Mammalian SWI/SNF Complex

We next turned to a complex of unknown structure and incompletely defined subunit composition, the mammalian SWI/SNF (mSWI/SNF or BAF) ATP-dependent chromatin remodeling complex. mSWI/SNF complexes are combinatorially assembled into 12–15 subunit heteromeric  $\sim 1.5$ –2 MDa complexes (Figure 4A), which utilize the energy of ATP hydrolysis to remodel nucleosomal architecture and oppose Polycomb repressive complexes, thus facilitating DNA accessibility and gene expression activation (Kadoch et al., 2017). Recent human genetic studies have unmasked recurrent mutations in the genes encoding mSWI/SNF subunits in over 20% of human cancers and in neurodevelopmental disorders such as intellectual disability syndromes (Kadoch and Crabtree, 2015). The mechanistic interpretation of the mSWI/SNF mutational spectrum is complicated by the incomplete functional characterization of several recently identified subunits, as well as the combinatorial subunit configurations produced by several paralogous, even tissue-specific, subunits (Kadoch et al., 2013). Therefore, we sought to apply our methodology to study the mSWI/SNF complex. The mSWI/SNF functional similarity network from both RNAi and CRISPR datasets revealed three distinct functional modules (Figures 4B, 4C, and S6A). The first corresponds to a core set of BAF complex components (ARID1A, SMARCB1, and SMARCE1 and the SMARCA4 ATPase subunit), while a second module is composed of distinguishing subunits of the PBAF variant of mSWI/SNF complexes (PBRM1, ARID2, and BRD7). The third functional module that did not correspond to any known configuration of the mSWI/SNF complex and is composed of one established mSWI/SNF subunit, SMARCD1; one recently discovered subunit, BRD9 (Kadoch et al., 2013); and one putative subunit, GLTSCR1 (Ho et al., 2009), whose highest ranked fitness correlations were with SMARCD1 and BRD9 (Figures 4B and S6A).

To experimentally determine whether these three functional modules exist as distinct biochemical entities, we performed size fractionation followed by immunoblot on nuclear extracts



**Figure 4. Fitness Correlation Mapping Identifies Biochemically Distinct Modules of Mammalian SWI/SNF Complexes**

(A) Schematic depicting subunits of the mammalian SWI/SNF family of ATP-dependent chromatin remodeling complexes.

(B) Fitness correlation network (from RNAi dataset) between mSWI/SNF subunits resolves three functional modules: core BAF (SMARCA4, ARID1A, SMARCB1, and SMARCE1), PBAF (PBRM1, ARID2, BRD7, and PHF10) and a novel functional module that contains two previously characterized subunits (SMARCD1 and BRD9) and one putative subunit (GLTSCR1).

(C) Hierarchical clustering performed on fitness profile correlations from the RNAi dataset groups subunits into distinct modules.

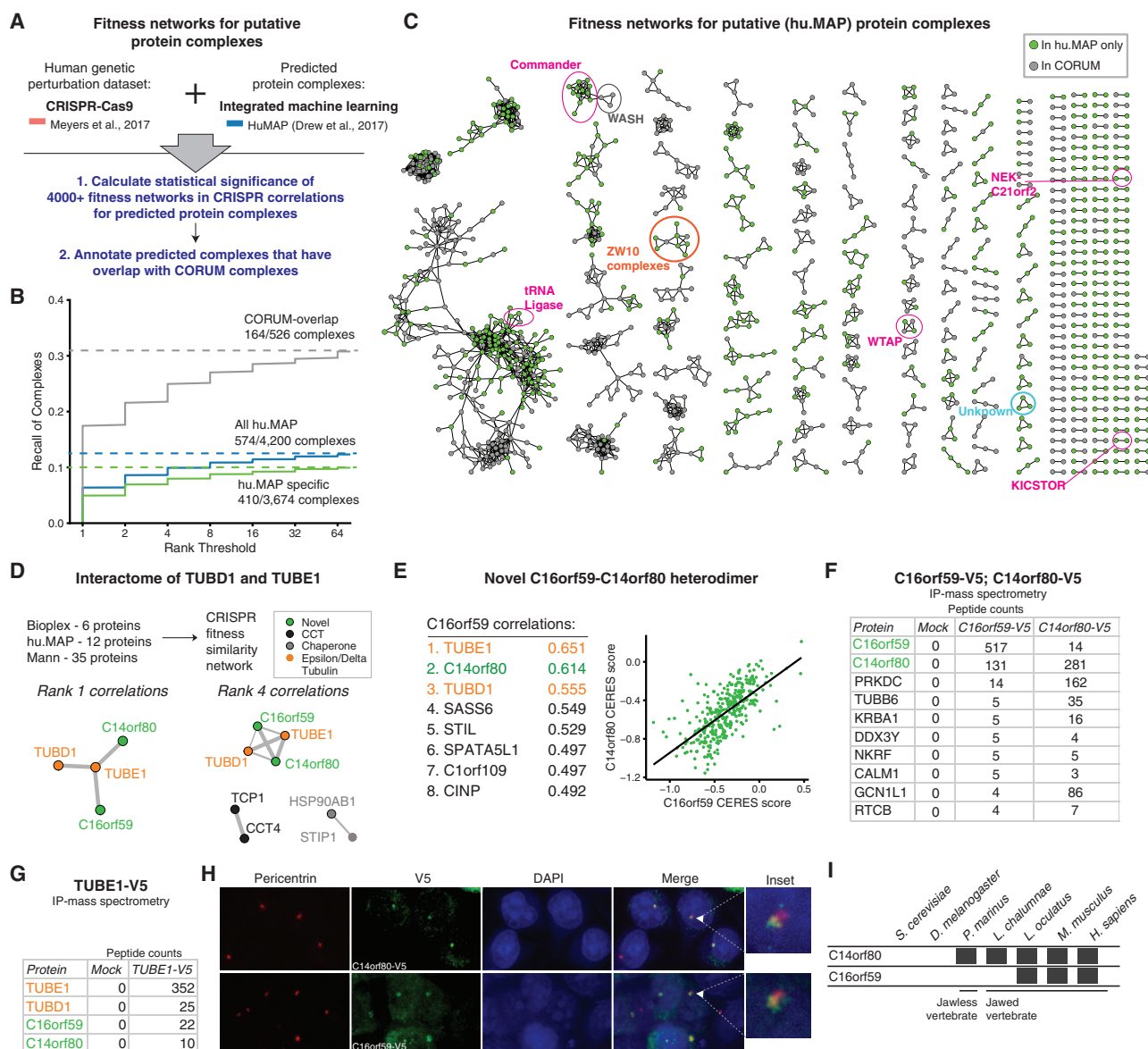
(D) Density sedimentation experiments using 10%–30% glycerol gradients performed on nuclear extracts from CCRF cells links two functional modules to known complexes, BAF (blue bar) and PBAF (red bar), and one to a novel assembly of distinct size and composition (green bar).

(E) Rare cancers characterized by mSWI/SNF perturbations exhibit mutually exclusive loss of one of the BAF core module genes or paralog families (containing SMARCA4, ARID1A, SMARCB1, and SMARCE1). SCCOHT, small cell carcinoma of the ovary, hypercalcemic type. In addition, specific intellectual disability syndromes are caused by heterozygous mutations in BAF core module genes.

isolated from CCRF cells. BAF- and PBAF-specific subunits separated into assemblies of different sizes, as shown by their migration in distinct fractions of 10%–30% glycerol gradients (Figure 4D). BRD9 migrated in lower-molecular-weight fractions of the gradient, indicating an unexpected smaller sub-assembly with SMARCD1. Immunoprecipitation studies further confirmed that the novel module binds the catalytic subunit SMARCA4 but fails to bind subunits found exclusively in other-sized fractions, such as ARID1A, ARID2, and BRD7 (Figure S6B). Immunoprecipitation of the SMARCA2 mSWI/SNF ATPase subunit from

cancer cell line nuclear extracts coupled with mass spectrometry-resolved high numbers of peptides of GLTSCR1 with minimal background signal (Figure S6C). These results are supported by large-scale published co-fractionation (Figure S6D) (Wan et al., 2015) and co-immunoprecipitation datasets (Figure S6E) (Huttlín et al., 2015), which suggest binding interactions but do not resolve mSWI/SNF modularity.

Together, the functional organization of mSWI/SNF subunits unveiled by functional similarity networks suggests the existence of three concurrently expressed mSWI/SNF family complexes



**Figure 5. A Combined Physical-Functional Interaction Map Highlights Validated and Novel Interactions**

(A) Strategy for the generation of fitness similarity networks for putative protein complexes. The statistical framework for identifying significant protein complex fitness correlation networks (Figure 2A) was applied to the hu.MAP complex dataset. Hu.MAP exhibits high level of complex enrichment within the CRISPR-Cas9 correlation dataset (Figure S1E).

(B) Fraction of hu.MAP protein complexes recalled in the CRISPR fitness correlation datasets. Of the 4,200 predicted complexes, 574 exhibit significant fitness networks.

(C) Statistically significant fitness correlation networks for hu.MAP complexes. Recently discovered protein complexes consisting of genes of unknown function are highlighted in magenta, and complexes with novel components that were selected for validation are labeled in orange and blue. Proteins found in the Core CORUM set are marked in gray, while proteins unique to the hu.MAP complex list are marked in green.

(D) To discover novel elements of the epsilon- and delta-tubulin interactome, 53 putative TUBE1 and TUBD1 interactors from three different large-scale protein-protein interaction networks were assembled and used to generate a fitness similarity network from the CRISPR-Cas9 dataset. Out of all 53 putative interactors, only two proteins, C16orf59 and C14orf80, exhibited top-ranked correlations with TUBE1 and TUBD1.

(E) Proteins exhibiting top-ranked fitness correlations with C16orf59 are predominantly centrosomal. One of the top-ranked correlations to C16orf59 is with another gene of unknown function, C14orf80. A scatterplot showing the correlation between CRISPR-Cas9 CERES scores of the C16orf59 and C14orf80 proteins across 300+ cell lines is shown.

(F) Immunoprecipitation (IP)/mass spectrometry results for V5-tagged C16orf59 and C14orf80 immunoprecipitations. Total peptide counts are indicated, ranked by overall abundance in the C16orf59 purification.

(G) IP/mass spectrometry of transiently transfected epsilon-tubulin (TUBE1) co-precipitates TUBD1 as well as the C16orf59-C14orf80 heterodimer.

(legend continued on next page)

that have distinct function and are assembled on a common catalytic subunit or module. Mechanistic dissection of this novel mSWI/SNF functional module on chromatin will require further study, particularly given recent advances in small molecule-based targeting of the BRD9 subunit (Hohmann et al., 2016). Importantly, functional SNF complexes were not/SNF complexes were not explained by co-expression; for example, *ARID1A* and *ARID2* assemble into mutually exclusive mSWI/SNF complexes (BAF and PBAF, respectively), are functionally distinct, and bind different sets of subunits, but exhibit one of the highest co-expression profiles among mSWI/SNF subunits in human normal tissue samples (Figures S6F and S6G). Finally, several rare diseases that are near-uniformly characterized by mSWI/SNF complex loss-of-function perturbations, spanning both rare cancer types and intellectual disability syndromes, contain homozygous or heterozygous mutations in genes which comprise the core BAF functional module (Figure 4E).

### Discovery of Novel Subunits and Protein Complexes from a Combined Physical-Functional Network Approach

Given the degree to which functional similarity networks recapitulate CORUM protein complex features, we conducted a second analysis on a set of predicted, non-validated protein complexes to nominate targets for further discovery. Based on the highly enriched overlap between the hu.MAP-predicted PPI network and the CRISPR functional similarity network at top ranks (Figure S1E), as well as the correlations in more lowly expressed, evolutionarily recent complexes (Figure 2E) captured in the CRISPR dataset, we used this dataset to identify significant fitness correlations among 4,000+ predicted hu.MAP protein complexes (Figure 5A). Our methodology identified 574 complexes recapitulated in the functional similarity network representing a recall of 14% at an FDR of 0.05 (Figure 5B). Of the 526 hu.MAP protein complexes showing >80% subunit overlap with annotated CORUM complexes, 164 protein complexes (31%) were recalled in the functional similarity network, compared with 410 of 3,674 protein complexes (11%) without CORUM overlap.

This set of 574 recalled complexes in the functional similarity network (Figure 5C; Table S6) is comprised of 1,962 total proteins, 1,478 of which do not appear in the Core CORUM dataset. Notably, we found that many of these fitness networks correspond to hu.MAP complexes that have been recently validated as novel protein complexes (i.e., C12orf66-SZT2 heteromer of the KICKSTOR complex [Wolfson et al., 2017] and C16orf62-COMMD [Phillips-Krawczak et al., 2015]) (Figures S7A and S7B). The recently discovered Commander complex (Wan et al., 2015) correlates with the WASHC4 and WASHC5 components of its known interacting complex, the WASH complex, and the KICKSTOR components SZT2 and C12orf66 correlate strongly with its mTOR pathway interactor, GATOR1 (Peng et al., 2017; Wolfson et al., 2017). The gene co-expression networks for the same complexes did not reveal these interactions (Figures S7C–S7F).

To identify novel interactions among these functional similarity networks for experimental validation, we scored each protein complex by taking the averaged products between the CRISPR correlation value and the hu.MAP probability weight over all gene pairs in the protein complex. We then ranked significant protein complexes by this score and identified complexes or subunits with no known literature annotation (Figure S8A and Table S6). Based on this strategy, we selected two unknown genes each with multiple interaction predictions for validation studies: C16orf59 (interaction pairs: TUBD1 and TUBE1), and C19orf25 (interaction pairs: NRZ complex members).

### C16orf59 and C14orf80 Form a Heterodimeric Complex that Binds Delta- and Epsilon-Tubulins

Tubulins are major elements of the eukaryotic cytoskeleton and are critical for cellular processes such as cell division and motility (Turk et al., 2015). Of the many characterized isoforms within the tubulin protein superfamily, the delta- and epsilon-tubulin variants remain incompletely understood. Recent work has suggested that these isoforms exist as an evolutionarily conserved module (Turk et al., 2015) and may be involved in forming triplet microtubules that are critical for centriole assembly (Wang et al., 2017a).

We assembled a combined TUBD1 and TUBE1 interactome consisting of 53 purported interactors from three different mass spectrometry datasets (Figures 5D and S8B). The CRISPR functional similarity network for these genes identified only 2 of the 53 purported interactors, and both were genes of no annotated function: C16orf59 as annotated above, and additionally, C14orf80 (Figure 5E). The fitness profile of C16orf59 also correlated strongly with SASS6, a core centrosome component that is necessary for centrosomal duplication (Leidel et al., 2005), suggesting that the function of C16orf59 is centrosome related (Figure 5E). Since one of the top fitness correlations of C16orf59 was C14orf80, a reported interactor in the Bioplex 2.0 network (Huttlin et al., 2017), we lentivirally introduced V5-tagged versions of C16orf59 and C14orf80 into HEK293T cells and performed immunoprecipitation with subsequent mass spectrometry-based proteomics. C16orf59 and C14orf80 reciprocally immunoprecipitated one another, suggesting that they form a heterodimeric protein complex (Figures 5F and S8C; Table S7). Both proteins precipitated with TUBE1 in a TUBE1-V5 immunoprecipitation (Figures 5G and S8C; Table S8), further supporting the evidence that the C16orf59 and C14orf80 heterodimer is a centrosomal interactor with delta- and epsilon-tubulin. To determine the subcellular localization of the C16orf59 and C14orf80 proteins, we performed immunofluorescence experiments in HEK293T cells containing V5-tagged C16orf59 and C14orf80 constructs. Consistent with the strong fitness correlation between C16orf59 and centrosome components, both C16orf59 and C14orf80 were found in the centrosomal components of nuclei marked by pericentrin as a centrosomal control (Figure 5H). Given the relatively recent evolutionary history of these two proteins—C14orf80 arose during jawless vertebrates, while C16orf59 arose in jawed

(H) Immunofluorescence performed for pericentrin (centrosomal marker) and V5 (C14orf80 and C16orf59), with DAPI nuclear stain. Both proteins exhibit centrosomal localization. Panel magnification: 60×.

(I) Evolutionary history of the C14orf80 and C16orf59 genes. Both are evolutionarily recent, with C16orf80 present only after the jawless-jawed vertebrate transition, while C14orf80 is present from jawless vertebrates forward.



vertebrates (Figure 5I)—our data suggest that the two uncharacterized proteins C16orf59 and C14orf80 form a vertebrate-specific centrosomal protein complex.

### **C19orf25 Selectively Binds the Cytoplasmic Module of the ZW10 Protein Complex Family**

The mammalian NRZ complex is composed of NBAS, ZW10, and RINT1 and is descended from its yeast predecessor, Dsl1. Both NRZ and Dsl1 complexes are involved in transport between the ER and the Golgi in their respective organisms (Tagaya et al., 2014). The ZW10 protein evolutionarily diverged to assemble into a second, nuclear-localized complex, RZZ, which facilitates dynein recruitment to the kinetochore (Vieugel et al., 2012). The RZZ complex is composed of four subunits: Rod (KNTC1 in humans), ZWILCH, and ZWINT, and the shared ZW10 subunit (Figure S9A). Previous efforts to purify ZW10 precipitated a factor, C19orf25, that did not appear to have kinetochore localization, despite binding ZW10 (Kops et al., 2005). We hypothesized that the C19orf25 functional similarity network would discern which form of the ZW10-nucleated complexes it associates with. The top fitness correlations of C19orf25 all are members of either NRZ, or the STX18 SNARE complex, which transiently docks NRZ on the ER membrane (STX18 and BNIP1) (Figures S9B and S9C). The “moonlighting” ZW10 protein is strongly correlated with both protein complex configurations, while C19orf25 forms specific correlations with the NRZ and STX18 complexes (Figures S9D and S9E). Consistent with the prediction that C19orf25 is functionally associated with NRZ and not RZZ, immunofluorescence experiments with V5-tagged C19orf25 confirmed cytoplasm-specific localization (Figure S9F). Finally, we were able to reproduce the previously observed interaction between C19orf25 and RINT1/ZW10 using immunoprecipitation/mass spectrometry of V5-C19orf25 from cytoplasmic extract (Figures S9G and S8C; Table S8). Given that C19orf25 is an evolutionarily recent protein complex subunit (Figure S9H), present only in bony vertebrates, our data suggest that C19orf25 is an evolutionarily recent addition to the cytoplasmic NRZ complex that arose concurrently with members of the RZZ complex.

## **DISCUSSION**

The ability to functionalize individual subunits and modules of protein complexes remains a major challenge, especially for those complexes with incompletely resolved protein subunit membership and structural information. Here, we demonstrate that protein complex componentry, as well as differential function between subunits, can be elucidated using large-scale genetic perturbation screens across diverse cellular contexts—in this case, hundreds of cancer cell lines. This study, to the best of our knowledge, represents the largest and most comprehensive analysis of protein complexes using fitness screening in human cells to date. These studies provide a conceptual framework for further study of functional relationships between proteins in both normal and disease-associated states (associated with genetic mutations, gene variants, or gene expression changes) as increasing fitness datasets continue to emerge.

Future work to merge disease genetics with physical and functional interactions may help reveal the molecular basis of certain human diseases. Indeed, examining disease-associated

alleles coding for interacting proteins may define convergent pathways and novel targets for therapeutic intervention. For instance, the putative hu.MAP-interacting subunits, C15orf41 and CDAN1, display strongly correlated CRISPR-Cas9 fitness profiles (Table S6), and the genes encoding these two proteins harbor mutually exclusive mutations in the majority of congenital dyserythropoietic anemias (Babbs et al., 2013). Similarly, our findings with respect to the core BAF functional module (SMARCA4, SMARCB1, SMARCE1, and ARID1A) are particularly timely and relevant to disease biology. Intriguingly, all rare cancer types known to be driven by mSWI/SNF complex perturbation (defined as  $\geq 70\%$  of tumors with protein-level loss of a single mSWI/SNF subunit) exhibit mutually exclusive and complete loss of one of the genes or paralog families in the core functional module of BAF complexes identified in our analysis (Figure 4E). The small percentage of these human tumor types not explained by their prevailing characteristic perturbation instead exhibit loss of one of the other members of the core BAF functional module we identified (Hasselblatt et al., 2011; Schneppenheimer et al., 2010; van den Munckhof et al., 2012). Finally, both Coffin-Siris and Nicolaides-Baraitser intellectual disability syndromes are driven by germline heterozygous mutations of core BAF module genes in a mutually exclusive manner (Figure 4E). Understanding the convergent functional contributions of mSWI/SNF subunits mutated in the specific cancers and intellectual disability syndromes highlighted above has remained a major recent challenge in the field. Our findings suggest a synergistic function between these four mSWI/SNF subunits, informing future studies to address the structural basis underlying these convergent functional correlations and disease-associated mutational patterns.

Several challenges still remain with respect to expanding the degree of protein complex capture from these or similar datasets, particularly for those complexes that do not demonstrate variable essentiality for cellular fitness. Additional approaches and screening readouts, such as cellular morphology (Rohban et al., 2017), may be able to better classify complex subunits that predominantly have morphological effects rather than fitness effects. This is particularly relevant for the utility of these datasets in the context of emerging genes-to-variants studies, and for the functional characterization of other human disease-linked genetic mutations. In addition, as similar genome-scale genetic perturbation screens are performed across increasingly larger and diverse sets of cell lines (and normal cell types), commensurate bioinformatic approaches will be required to address normalization methods, and to enable further integration with machine learning-based PPI classifiers and other ensemble approaches. We provide all fitness correlations from both RNAi and CRISPR-Cas9 datasets as well as the statistically significant fitness networks for both CORUM and hu.MAP complexes as resources for the larger research community.

## **STAR★METHODS**

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **CONTACT FOR REAGENT AND RESOURCE SHARING**



## ● EXPERIMENTAL MODEL AND SUBJECT DETAILS

### ● METHOD DETAILS

- Lentiviral Packaging and Infection
- Nuclear Extract Isolation, Immunoprecipitation, and Mass-Spectrometry of BAF Complex Subunits
- Density Sedimentation (10-30% Glycerol Gradients)
- Immunoblot
- Antibodies Used
- Large-Scale Purification of Factors for Mass Spectrometry (C14orf80, C16orf59, C19orf25, TUBE1)
- Immunofluorescence (IF)
- Imaging and Analysis

### ● QUANTIFICATION AND STATISTICAL ANALYSIS

- Using Fitness Thresholds to Define Genes for Analysis
- Fitness Profile Plots
- Fitness Variation
- Enrichment of Correlated Gene Pairs among Protein-Protein Interactions
- Similarity Network Significance for Protein Complexes
- Network Visualizations
- Interface Size Dataset
- Modularity Overlap Enrichment
- Heatmaps and Clustering
- Scoring Significant hu.MAP Complexes
- Gene Evolution Tables

### ● DATA AND SOFTWARE AVAILABILITY

## SUPPLEMENTAL INFORMATION

Supplemental Information includes nine figures, eight tables, and two data files can be found with this article online at <https://doi.org/10.1016/j.cels.2018.04.011>.

## ACKNOWLEDGMENTS

This work was supported by the National Science Foundation Graduate Research Fellowship (2015185722), the NIH Training Grant in Genetics and Genomics (T32GM096911), and the Quantitative Cell Biology Network (NSF MCB-1411898) to J.P. C.K. is supported by the NIH DP2 Director's New Innovator Award (1DP2CA195762-01), the American Cancer Society Research Scholar Award RSG-14-051-01-DMC, and the Pew-Stewart Scholars in Cancer Research Grant. J.M. is supported by a Medical Research Council Career Development Award (MR/M02122X/1). The GPP at the Broad Institute is supported by the Carlos Slim Foundation in Mexico through the Slim Initiative for Genomic Medicine and NCI grant U01 CA176058 to W.C.H. The project was also supported by award number T32GM007753 from the National Institute of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. We thank members of the Kadoch Lab, the Genetics Perturbation Platform (GPP) at the Broad Institute, and B. Tye, M. Sonnett, R. Gopalakrishnan, F. Winston, R.E. Kingston, M.L. Meyerson, and S.A. Teichmann for helpful discussions.

## AUTHOR CONTRIBUTIONS

J.P., R.M.M., A.T., and C.K. conceived of and designed the study. J.P., R.M.M., A.E.S., and A.T. performed all analyses. J.P., B.C.M., and N.M. performed all experimental validation studies. S.H.C. contributed novel insights and data. F.V., B.A.W., W.C.H., and A.T. performed and directed RNAi- and CRISPR-Cas9-based screening efforts as part of Project Achilles (Broad Institute), J.N.W. and J.A.M. provided structural data and performed structural analyses of protein complexes. J.P., R.M.M., and C.K. wrote the manuscript.

## DECLARATION OF INTERESTS

C.K. is a scientific founder, shareholder, and consultant of Foghorn Therapeutics, Inc. (Cambridge, MA).

Received: February 26, 2018

Revised: April 3, 2018

Accepted: April 24, 2018

Published: May 16, 2018

## REFERENCES

- Aguirre, A.J., Meyers, R.M., Weir, B.A., Vazquez, F., Zhang, C.Z., Ben-David, U., Cook, A., Ha, G., Harrington, W.F., Doshi, M.B., et al. (2016). Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discov.* 6, 914–929.
- Ahnert, S.E., Marsh, J.A., Hernandez, H., Robinson, C.V., and Teichmann, S.A. (2015). Principles of assembly reveal a periodic table of protein complexes. *Science* 350, aaa2245.
- Babbs, C., Roberts, N.A., Sanchez-Pulido, L., McGowan, S.J., Ahmed, M.R., Brown, J.M., Sabry, M.A., Bentley, D.R., McVean, G.A., Donnelly, P., et al. (2013). Homozygous mutations in a predicted endonuclease are a novel cause of congenital dyserythropoietic anemia type I. *Haematologica* 98, 1383–1387.
- Baliga, N.S., Björkegren, J., Boeke, J.D., Boutros, M., Crawford, N., Dudley, A.M., Farber, C.R., Jones, A., Levey, A.I., Lusi, A.J., et al. (2017). The state of systems genetics in 2017. *Cell Syst.* 4, 7–15.
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607.
- Baryshnikova, A., Costanzo, M., Myers, C.L., Andrews, B., and Boone, C. (2013). Genetic interaction networks: toward an understanding of heritability. *Annu. Rev. Genomics Hum. Genet.* 14, 111–133.
- Bertomeu, T., Coulombe-Huntington, J., Chatr-Aryamontri, A., Bourdages, K.G., Coyaoud, E., Raught, B., Xia, Y., and Tyers, M. (2018). A high-resolution genome-wide CRISPR/Cas9 viability screen reveals structural features and contextual diversity of the human cell-essential proteome. *Mol. Cell. Biol.* 38, <https://doi.org/10.1128/MCB.00302-17>.
- Blomen, V.A., Májek, P., Jae, L.T., Bigenzahn, J.W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F.R., Olk, N., Stukalov, A., et al. (2015). Gene essentiality and synthetic lethality in haploid human cells. *Science* 350, 1092–1096.
- Boettcher, M., Tian, R., Blau, J.A., Markegard, E., Wagner, R.T., Wu, D., Mo, X., Biton, A., Zaitlen, N., Fu, H., et al. (2018). Dual gene activation and knockout screen reveals directional dependencies in genetic networks. *Nat. Biotechnol.* 36, 170.
- Carter, R., and Drouin, G. (2009). The increase in the number of subunits in eukaryotic RNA polymerase III relative to RNA polymerase II is due to the permanent recruitment of general transcription factors. *Mol. Biol. Evol.* 27, 1035–1043.
- Collins, S.R., Miller, K.M., Maas, N.L., Roguev, A., Fillingham, J., Chu, C.S., Schuldiner, M., Gebbia, M., Recht, J., Shales, M., et al. (2007). Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* 446, 806–810.
- Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L., Toufighi, K., Mostafavi, S., et al. (2010). The genetic landscape of a cell. *Science* 327, 425–431.
- Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S.D., et al. (2016). A global genetic interaction network maps a wiring diagram of cellular function. *Science* 353, <https://doi.org/10.1126/science.aaf1420>.
- Cowley, G.S., Weir, B.A., Vazquez, F., Tamayo, P., Scott, J.A., Rusin, S., East-Seletsky, A., Ali, L.D., Gerath, W.F., Pantel, S.E., et al. (2014). Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies. *Sci. Data* 1, 140035.

- Dambacher, C.M., Worden, E.J., Herzik, M.A., Martin, A., and Lander, G.C. (2016). Atomic structure of the 26S proteasome lid reveals the mechanism of deubiquitinase inhibition. *Elife* 5, e13027.
- de Juan, D., Pazos, F., and Valencia, A. (2013). Emerging methods in protein co-evolution. *Nat. Rev. Genet.* 14, 249–261.
- Doench, J.G. (2018). Am I ready for CRISPR? A user's guide to genetic screens. *Nat. Rev. Genet.* 19, 67–80.
- Drew, K., Lee, C., Huizar, R.L., Tu, F., Borgeson, B., McWhite, C.D., Ma, Y., Wallingford, J.B., and Marcotte, E.M. (2017). Integration of over 9,000 mass spectrometry experiments builds a global map of human protein complexes. *Mol. Syst. Biol.* 13, 932.
- Du, D., Roguev, A., Gordon, D.E., Chen, M., Chen, S.-H.H., Shales, M., Shen, J.P., Ideker, T., Mali, P., Qi, L.S., et al. (2017). Genetic interaction mapping in mammalian cells using CRISPR interference. *Nat. Methods* 14, 577–580.
- Gingras, A.-C., Gstaiger, M., Raught, B., and Aebersold, R. (2007). Analysis of protein complexes using mass spectrometry. *Nat. Rev. Mol. Cell Biol.* 8, 645–654.
- Han, K., Jeng, E.E., Hess, G.T., Morgens, D.W., Li, A., and Bassik, M.C. (2017). Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat. Biotechnol.* 35, 463–474.
- Hart, T., Chandrashekar, M., Aregger, M., Steinhart, Z., Brown, Kevin R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., et al. (2015). High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* 163, 1515–1526.
- Hart, T., Koh, C., and Moffat, J. (2017a). Coessentiality and cofunctionality: a network approach to learning genetic vulnerabilities from cancer cell line fitness screens. *bioRxiv*. <https://doi.org/10.1101/134346>.
- Hart, T., Tong, A.H.Y.H.Y., Chan, K., Van Leeuwen, J., Seetharaman, A., Aregger, M., Chandrashekar, M., Hustedt, N., Seth, S., Noonan, A., et al. (2017b). Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. *G3 (Bethesda)* 7, 2719–2727.
- Hasselblatt, M., Gesk, S., Oyen, F., Rossi, S., Viscardi, E., Giangaspero, F., Giannini, C., Judkins, A.R., Frühwald, M.C., Obser, T., et al. (2011). Nonsense mutation and inactivation of SMARCA4 (BRG1) in an atypical teratoid/rhabdoid tumor showing retained SMARCB1 (INI1) expression. *Am. J. Surg. Pathol.* 35, 933–935.
- Hein, M.Y., Hubner, N.C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I.A., Weisswange, I., Mansfeld, J., Buchholz, F., et al. (2015). A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* 163, 712–723.
- Ho, L., Ronan, J.L., Wu, J., Staahl, B.T., Chen, L., Kuo, A., Lessard, J., Nesvizhskii, A.I., Ranish, J., and Crabtree, G.R. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc. Natl. Acad. Sci. USA* 106, 5181–5186.
- Hohmann, A.F., Martin, L.J., Minder, J.L., Roe, J.-S.S., Shi, J., Steurer, S., Bader, G., McConnell, D., Pearson, M., Gerstberger, T., et al. (2016). Sensitivity and engineered resistance of myeloid leukemia cells to BRD9 inhibition. *Nat. Chem. Biol.* 12, 672–679.
- Huttlin, E.L., Ting, L., Bruckner, R.J., Gebreab, F., Gygi, M.P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., et al. (2015). The BioPlex network: a systematic exploration of the human interactome. *Cell* 162, 425–440.
- Huttlin, E.L., Bruckner, R.J., Paulo, J.A., Cannon, J.R., Ting, L., Baltier, K., Colby, G., Gebreab, F., Gygi, M.P., Parzen, H., et al. (2017). Architecture of the human interactome defines protein communities and disease networks. *Nature* 545, 505–509.
- Jansen, R., Greenbaum, D., and Gerstein, M. (2002). Relating whole-genome expression data with protein-protein interactions. *Genome Res.* 12, 37–46.
- Jeronimo, C., Langelier, M.-F.F., Bataille, A.R., Pascal, J.M., Pugh, B.F., and Robert, F. (2016). Tail and kinase modules differently regulate core mediator recruitment and function in vivo. *Mol. Cell* 64, 455–466.
- Kadoch, C., and Crabtree, G.R. (2015). Mammalian SWI/SNF chromatin remodeling complexes and cancer: mechanistic insights gained from human genomics. *Sci. Adv.* 1, e1500447.
- Kadoch, C., Hargreaves, D.C., Hodges, C., Elias, L., Ho, L., Ranish, J., and Crabtree, G.R. (2013). Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* 45, 592–601.
- Kadoch, C., Williams, R.T., Calarco, J.P., Miller, E.L., Weber, C.M., Braun, S.M., Pulice, J.L., Chory, E.J., and Crabtree, G.R. (2017). Dynamics of BAF-Polycomb complex opposition on heterochromatin in normal and oncogenic states. *Nat. Genet.* 49, 213–222.
- Kelley, R., and Ideker, T. (2005). Systematic interpretation of genetic interactions using protein networks. *Nat. Biotechnol.* 23, 561–566.
- Kops, G., Kim, Y., Weaver, B.A.A., Mao, Y., McLeod, I., Yates, J.R., Tagaya, M., and Cleveland, D.W. (2005). ZW10 links mitotic checkpoint signaling to the structural kinetochore. *J. Cell Biol.* 169, 49–60.
- Leidel, S., Delattre, M., Cerutti, L., Baumer, K., and Gönczy, P. (2005). SAS-6 defines a protein family required for centrosome duplication in *C. elegans* and in human cells. *Nat. Cell Biol.* 7, 115–125.
- Mann, C., Buhler, J.M., Treich, I., and Sentenac, A. (1987). RPC40, a unique gene for a subunit shared between yeast RNA polymerases A and C. *Cell* 48, 627–637.
- McDonald, E.R., de Weck, A., Schlabach, M.R., Billy, E., Mavrakis, K.J., Hoffman, G.R., Belur, D., Castelletti, D., Frias, E., Gampa, K., et al. (2017). Project DRIVE: a compendium of cancer dependencies and synthetic lethal relationships uncovered by large-scale, deep RNAi screening. *Cell* 170, 577–592.e10.
- Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N.V., Montgomery, P.G., Cowley, G.S., Pantel, S., et al. (2017). Computational correction of copy-number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* 49, 1779–1784.
- Najm, F.J., Strand, C., Donovan, K.F., Hegde, M., Sanson, K.R., Vaimberg, E.W., Sullender, M.E., Hartenian, E., Kalani, Z., Fusi, N., et al. (2018). Orthologous CRISPR-Cas9 enzymes for combinatorial genetic screens. *Nat. Biotechnol.* 36, 179–189.
- Nozawa, K., Schneider, T.R., and Cramer, P. (2017). Core Mediator structure at 3.4 Å extends model of transcription initiation complex. *Nature* 545, 248–251.
- Okamura, Y., Aoki, Y., Obayashi, T., Tadaka, S., Ito, S., Narise, T., and Kinoshita, K. (2015). COXPRESdb in 2015: coexpression database for animal species by DNA-microarray and RNAseq-based expression data with multiple quality assessment systems. *Nucleic Acids Res.* 43, 6.
- Pan, X., Yuan, D.S., Xiang, D., Wang, X., Sookhai-Mahadeo, S., Bader, J.S., Hieter, P., Spencer, F., and Boeke, J.D. (2004). A robust toolkit for functional profiling of the yeast genome. *Mol. Cell* 16, 487–496.
- Peng, M., Yin, N., and Li, M.O. (2017). SZT2 dictates GATOR control of mTORC1 signalling. *Nature* 543, 433–437.
- Pereira-Leal, J.B., Levy, E.D., and Teichmann, S.A. (2006). The origins and evolution of functional modules: lessons from protein complexes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 507–517.
- Petrenko, N., Jin, Y., Wong, K.H., and Struhl, K. (2016). Mediator undergoes a compositional change during transcriptional activation. *Mol. Cell* 64, 443–454.
- Phillips-Krawczak, C.A., Singla, A., Starokadomskyy, P., Deng, Z., Osborne, D.G., Li, H., Dick, C.J., Gomez, T.S., Koenecke, M., Zhang, J.-S.S., et al. (2015). COMMD1 is linked to the WASH complex and regulates endosomal trafficking of the copper transporter ATP7A. *Mol. Biol. Cell* 26, 91–103.
- Ramani, A.K., Li, Z., Hart, G.T., Carlson, M.W., Boutz, D.R., and Marcotte, E.M. (2008). A map of human protein interactions derived from co-expression of human mRNAs and their orthologs. *Mol. Syst. Biol.* 4, 180.
- Rohban, M.H., Singh, S., Wu, X., Berthet, J.B., Bray, M.-A.A., Shrestha, Y., Varelas, X., Boehm, J.S., and Carpenter, A.E. (2017). Systematic morphological profiling of human gene and allele function via cell painting. *eLife* 6, <https://doi.org/10.7554/eLife.24060>.
- Rolland, T., Tazan, M., Charlotiaux, B., Pevzner, S.J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R., et al. (2014). A proteome-scale map of the human interactome network. *Cell* 159, 1212–1226.
- Ruepp, A., Waegle, B., Lechner, M., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Mewes, H.W. (2010). CORUM:

the comprehensive resource of mammalian protein complexes—2009. *Nucleic Acids Res.* 38, D497–D501.

Schneppenheimer, R., Frühwald, M.C., Gesk, S., Hasselblatt, M., Jeibmann, A., Kordes, U., Kreuz, M., Leuschner, I., Martin Subero, J.I., Obser, T., et al. (2010). Germline nonsense mutation and somatic inactivation of SMARCA4/BRG1 in a family with rhabdoid tumor predisposition syndrome. *Am. J. Hum. Genet.* 86, 279–284.

Schuldiner, M., Collins, S.R., Thompson, N.J., Denic, V., Bhamidipati, A., Punna, T., Ihmels, J., Andrews, B., Boone, C., Greenblatt, J.F., et al. (2005). Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* 123, 507–519.

Shen, J.P., Zhao, D., Sasik, R., Luebeck, J., Birmingham, A., Bojorquez-Gomez, A., Licon, K., Klepper, K., Pekin, D., Beckett, A.N., et al. (2017). Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. *Nat. Methods* 14, 573–576.

Smedley, D., Haider, S., Durinck, S., Pandini, L., Provero, P., Allen, J., Arnaiz, O., Awedh, M., Baldock, R., Barbiera, G., et al. (2015). The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res.* 43, W589–W598.

Spedale, G., Timmers, H.T., and Pijnappel, W.W. (2012). ATAC-king the complexity of SAGA during evolution. *Genes Dev.* 26, 527–541.

Tagaya, M., Arasaki, K., Inoue, H., and Kimura, H. (2014). Moonlighting functions of the NRZ (mammalian Dsl1) complex. *Front. Cell Dev. Biol.* 2, 25.

Tan, Q., Prysak, M.H., and Woychik, N.A. (2003). Loss of the Rpb4/Rpb7 sub-complex in a mutant form of the Rpb6 subunit shared by RNA polymerases I, II, and III. *Mol. Cell. Biol.* 23, 3329–3338.

Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L.M., et al. (2017). A subcellular map of the human proteome. *Science* 356, <https://doi.org/10.1126/science.aal3321>.

Tong, A., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M., et al. (2004). Global mapping of the yeast genetic interaction network. *Science* 303, 808–813.

Tsai, K.-L.L., Yu, X., Gopalan, S., Chao, T.-C.C., Zhang, Y., Florens, L., Washburn, M.P., Murakami, K., Conaway, R.C., Conaway, J.W., et al. (2017). Mediator structure and rearrangements required for holoenzyme formation. *Nature* 544, 196–201.

Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a cancer dependency map. *Cell* 170, 564–576.e16.

Turk, E., Wills, A.A., Kwon, T., Sedzinski, J., Wallingford, J.B., and Stearns, T. (2015). Zeta-tubulin is a member of a conserved tubulin module and is a component of the centriolar basal foot in multiciliated cells. *Curr. Biol.* 25, 2177–2183.

van den Munckhof, P., Christiaans, I., Kenter, S.B., Baas, F., and Hulsebos, T.J. (2012). Germline SMARCB1 mutation predisposes to multiple meningiomas and schwannomas with preferential location of cranial meningiomas at the falx cerebri. *Neurogenetics* 13, 1–7.

Vleugel, M., Hoogendoorn, E., Snel, B., and Kops, G.J. (2012). Evolution and function of the mitotic checkpoint. *Dev. Cell* 23, 239–250.

Wan, C., Borgeson, B., Phanse, S., Tu, F., Drew, K., Clark, G., Xiong, X., Kagan, O., Kwan, J., Bezginov, A., et al. (2015). Panorama of ancient metazoan macromolecular complexes. *Nature* 525, 339–344.

Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S., and Sabatini, D.M. (2015). Identification and characterization of essential genes in the human genome. *Science* 350, 1096–1101.

Wang, J.T., Kong, D., Hoerner, C.R., Loncarek, J., and Stearns, T. (2017a). Centriole triplet microtubules are required for stable centriole formation and inheritance in human cells. *Elife* 6, <https://doi.org/10.7554/eLife.29061>.

Wang, T., Yu, H., Hughes, N.W., Liu, B., Kendirli, A., Klein, K., Chen, W.W., Lander, E.S., and Sabatini, D.M. (2017b). Gene essentiality profiling reveals gene networks and synthetic lethal interactions with oncogenic Ras. *Cell* 168, 890–942505984.

Werner, F., and Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* 9, 85–98.

Wild, T., and Cramer, P. (2012). Biogenesis of multisubunit RNA polymerases. *Trends Biochem. Sci.* 37, 99–105.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* 67, 235–242.

Wolfson, R.L., Chantranupong, L., Wyant, G.A., Gu, X., Orozco, J.M., Shen, K., Condon, K.J., Petri, S., Kadir, J., Scaria, S.M., et al. (2017). KICSTOR recruits GATOR1 to the lysosome and is necessary for nutrients to regulate mTORC1. *Nature* 543, 438–442.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
SMARCA4 Brg-1 (G-7) antibody	Santa Cruz	Cat#sc-17796; RRID: AB_626762
ARID1A/BAF250 Antibody	Bethyl	Cat#A301-041; RRID: AB_2060365
SMARCE1-human antibody	Bethyl	Cat#A300-810A; RRID: AB_577243
SMARCB1 Anti-Ini1 Monoclonal Antibody	Santa Cruz	Cat#sc-166165; RRID: AB_2270651
SS18	Cell Signaling	Cat#21792; RRID: AB_2728667
DPF2	Abcam	Cat#ab134942; RRID: AB_2728668
SMARCD2 (F-34)	Santa Cruz	Cat#sc-101162; RRID: AB_1129531
SMARCC1/ BAF155 (H-76)	Santa Cruz	Cat#sc-10756; RRID: AB_2191997
PBRM1/BAF180 Antibody	Bethyl	Cat#A301-591A ; RRID: AB_1078808
ARID2/BAF200 Monoclonal (E-3)	Santa Cruz	Cat#sc-166117; RRID: AB_2060382
BRD7 (B-8)	Santa Cruz	Cat#sc-376180; RRID: AB_10989389
SMARCD1 / BAF60a (23)	Santa Cruz	Cat#sc-135843; RRID: AB_2192137
BRD9	Abcam	Cat#ab66443; RRID: AB_2066035
PHF10	Invitrogen	Cat#PA5-30678; RRID: AB_2548152
SMARCA4 (IP)	Abcam	Cat#EPNCIR111A; RRID: AB_10861578
Pericentrin	Abcam	Cat#ab4448; RRID: AB_304461
V5 (mouse)	Thermo	Cat#R960-25; RRID: AB_2556564
V5 (rabbit)	Cell Signaling	Cat#13202; RRID: AB_2687461
Anti-KDEL	Abcam	Cat#ab50601; RRID: AB_880636
Goat anti-Mouse IgG Alexa Fluor Plus 555	Thermo Fisher	Cat#A32727; RRID: AB_2633276
Goat anti-Rat IgG Alexa Fluor 647	Thermo Fisher	Cat#A-21247; RRID: AB_141778
Goat anti-Rabbit IgG Alexa Fluor 546	Cell Signaling	Cat#A-11010; RRID: AB_143156
Donkey Anti-Rabbit IgG (H+L) Polyclonal Antibody, Alexa Fluor 647 Conjugated	Thermo Fisher	RRID: AB_2536183
Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Thermo Fisher	RRID: AB_2534069
<b>Deposited Data</b>		
Project Achilles: RNAi	Tsherniak et al., 2017	<a href="https://portals.broadinstitute.org/achilles/datasets/all">https://portals.broadinstitute.org/achilles/datasets/all</a>
Project Achilles: CRISPR-Cas9 Avana	Meyers et al., 2017	<a href="https://portals.broadinstitute.org/achilles/datasets/all">https://portals.broadinstitute.org/achilles/datasets/all</a>
Project Achilles: CRISPR-Cas9 GeCKO (3.3.8)	Aguirre et al., 2016	<a href="https://portals.broadinstitute.org/achilles/datasets/all">https://portals.broadinstitute.org/achilles/datasets/all</a>
Project Achilles: RNAi 54k library ATARIS (2.4.3)	Cowley et al., 2014	<a href="https://portals.broadinstitute.org/achilles/datasets/all">https://portals.broadinstitute.org/achilles/datasets/all</a>
AML CRISPR screens	Wang et al., 2017b	<a href="http://sabatinilab.wi.mit.edu/wang/2017">sabatinilab.wi.mit.edu/wang/2017</a>
BioPlex 2.0 Interaction List	Huttlin et al., 2017	<a href="http://bioplex.hms.harvard.edu/downloadInteractions.php">http://bioplex.hms.harvard.edu/downloadInteractions.php</a> ; <b>RRID:SCR_016144</b>
Mann PPI List	Hein et al., 2015	<a href="http://www.sciencedirect.com/science/article/pii/S0092867415012702?via%3Dihub">http://www.sciencedirect.com/science/article/pii/S0092867415012702?via%3Dihub</a>
Hu.MAP complex data	Drew et al., 2017	<a href="http://msb.embopress.org/content/13/6/932">http://msb.embopress.org/content/13/6/932</a>
Marcotte complex data	Wan et al., 2015	<a href="http://www.nature.com/nature/journal/v525/n7569/full/nature14877.html">http://www.nature.com/nature/journal/v525/n7569/full/nature14877.html</a>
Yeast 2 Hybrid Interactome (HURI)	Rolland et al., 2014	<a href="http://interactome.dfci.harvard.edu/H_sapiens/download/HI-II-14.tsv">http://interactome.dfci.harvard.edu/H_sapiens/download/HI-II-14.tsv</a> ; <b>RRID:SCR_015670</b>

(Continued on next page)

### Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Subcellular localization interactome	Thul et al., 2017	<a href="http://science.sciencemag.org/content/356/6340/eaal3321">http://science.sciencemag.org/content/356/6340/eaal3321</a>
CORUM Core complexes (3.5.2017 release)	Ruepp et al., 2010	<a href="http://mips.helmholtz-muenchen.de/corum/#download">http://mips.helmholtz-muenchen.de/corum/#download</a> ; <b>RRID:SCR_002254</b>
AREAIMOL Buried surface area for mammalian structures	This paper	N/A
CCLE RNA-seq	Broad Institute/NCI	<a href="https://ocg.cancer.gov/ctd2-data-project/translational-genomics-research-institute-quantified-cancer-cell-line-encyclopedia">https://ocg.cancer.gov/ctd2-data-project/translational-genomics-research-institute-quantified-cancer-cell-line-encyclopedia</a> ; <b>RRID:SCR_013836</b>
COXPRESdb (Hsa3.c1-0)	Okamura et al., 2015	<a href="http://coxpresdb.jp/download.shtml">http://coxpresdb.jp/download.shtml</a>
Yeast Protein Complexes	Costanzo et al., 2016	<a href="http://science.sciencemag.org/content/353/6306/aaf1420">http://science.sciencemag.org/content/353/6306/aaf1420</a>
Yeast Genetic interaction data	Costanzo et al., 2016	<a href="http://thecellmap.org/costanzo2016/">http://thecellmap.org/costanzo2016/</a>
Human – Mouse gene dN/dS	BioMart	<a href="http://useast.ensembl.org/biomart/martview/">http://useast.ensembl.org/biomart/martview/</a> ; <b>RRID:SCR_002987</b>
Experimental Models: Cell Lines		
293T	ATCC	ATCC #CRL-3216
CCRF	ATCC	ATCC #CCL-119
A549	ATCC	ATCC #CCL-185
Recombinant DNA		
pLX_TRC317 TUBE1	Broad GPP	TRCN0000479747
pLX_TRC317 C16orf59	Broad GPP	TRCN0000481386
pLX_TRC317 C14orf80	Broad GPP	TRCN0000472525
pLX_TRC317 C19orf25	Broad GPP	TRCN0000473718
psPAX2	Addgene	<a href="https://www.addgene.org/12260/">https://www.addgene.org/12260/</a>
pMD2.G	Addgene	<a href="https://www.addgene.org/12259/">https://www.addgene.org/12259/</a>
Software and Algorithms		
Igraph		<a href="http://igraph.org">http://igraph.org</a>
ggiraph	David Gohel	<a href="http://davidgohel.github.io/ggiraph">http://davidgohel.github.io/ggiraph</a>
Ggraph	Thomas Lin Pedersen	<a href="https://github.com/thomasp85/ggraph">https://github.com/thomasp85/ggraph</a>
Ggjoy	Claus O. Wilke	<a href="https://cran.r-project.org/web/packages/ggjoy/index.html">https://cran.r-project.org/web/packages/ggjoy/index.html</a>
Cytoscape	Institute for Systems Biology	<a href="http://cytoscape.org">http://cytoscape.org</a> ; <b>RRID:SCR_003032</b>
Computational Project No. 4: Software for Macromolecular X-Ray Crystallography		<a href="http://www.ccp4.ac.uk/">http://www.ccp4.ac.uk/</a> ; <b>RRID:SCR_007255</b>
Gplots		<a href="https://cran.r-project.org/web/packages/gplots/index.html">https://cran.r-project.org/web/packages/gplots/index.html</a>
ImageJ	NIH	<a href="https://imagej.nih.gov/ij/">https://imagej.nih.gov/ij/</a> ; <b>RRID:SCR_003070</b>
Other		
Polybrene transfection reagent	Millipore Sigma	TR-1003-G
Gammabind G Sepharose beads	GE	17088501
14 x 89 mm polyallomer centrifuge tube	Beckman Coulter	331327
NuPAGE 4-12% Bis-Tris Protein Gels, 1.0 mm, 12-well	Thermo Fisher Scientific	NP0322BOX

### CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Cigall Kadoch, Ph.D. ([cigall\\_kadoch@dfci.harvard.edu](mailto:cigall_kadoch@dfci.harvard.edu)).



## EXPERIMENTAL MODEL AND SUBJECT DETAILS

293T (ATCC #CRL-3216) and A549 (ATCC #CCL-185, male) cell lines were passaged and grown in DMEM media, supplemented with Glutamax and pen/strep and 10% FBS. CCRF-CEM (ATCC #CCL-119, female) cells were grown in suspension in RPMI 1640 supplemented with 10% FBS, glutamine, and pen/strep.

## METHOD DETAILS

### Lentiviral Packaging and Infection

pMD2.G and psPAX2 lentiviral packaging vectors was co-transfected with pLX317 vector containing the clone of interest into HEK-293T cells, using polyethylenimine as a transfection reagent. Cells were incubated for 72 hours, and the media was filtered with a 0.4  $\mu$ m filter before being either concentrated with an ultracentrifuge (20,000 RPM for 2.5 hours) or added directly to cells plated at 70% confluence with 1:1000 Polybrene.

### Nuclear Extract Isolation, Immunoprecipitation, and Mass-Spectrometry of BAF Complex Subunits

Harvested cells were incubated in Buffer A (25 mM HEPES pH 7.6, 5 mM MgCl<sub>2</sub>, 25 mM KCl, 0.05 mM EDTA, 10% glycerol and 0.1% NP40 with protease inhibitor (Roche), 1 mM DTT and 1 mM phenylmethylsulfonyl fluoride (PMSF)) for 10 minutes and the pellets were resuspended in 600  $\mu$ l of Buffer C (10 mM HEPES pH 7.6, 3 mM MgCl<sub>2</sub>, 100 mM KCl, 0.5 mM EDTA and 10% glycerol with protease inhibitor, 1 mM DTT and 1 mM PMSF) with 67  $\mu$ l of 3 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> for 20 minutes. The lysates were spun down using a tabletop ultracentrifuge at 100,000 rpm at 4°C for 10 minutes. Nuclear extracts were precipitated with 200 mg of (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> on ice for 20 minutes and finally purified as pellets by ultracentrifugation at 100,000 rpm at 4°C for 10 minutes. The pellets were resuspended in IP Buffer (150 mM NaCl, 50 mM Tris-HCl pH 7.5, 1 mM EDTA and 1% Triton-X100 with protease inhibitor, 1 mM DTT and 1 mM PMSF) for the subsequent experiments. Immunoprecipitation was performed with antibodies targeting BRD7 (Bethyl, A302-304A), BRD9 (Abcam, ab137245), ARID1A (Bethyl, A301-041), SMARCA4 (Abcam, [EPNCIR111A], ab110641) and SMARCA2 (Bethyl, A301-015A).

For proteomic analysis, antibodies were crosslinked with dimethyl pimelimidate (DMP) to Gammabind G Sepharose beads (GE) prior to immunoprecipitation from nuclear extract. Captured protein was eluted using 6M urea/600 mM NaCl and digested with trypsin. Mass spectrometry was performed (using a Thermo Exactive Plus Orbitrap) by the Taplin Mass Spec Facility (Harvard Medical School). To analyze the results, peptides that were present in control precipitations (mock) were removed from the bait peptide list, and each protein was then ranked according to the number of total peptides captured.

### Density Sedimentation (10–30% Glycerol Gradients)

Nuclear extract (800  $\mu$ g, quantified by Bradford assay) was resuspended in 200  $\mu$ l of 0% glycerol HEMG buffer (supplemented with protease inhibitors and DTT) and overlaid onto a 11 ml 10%–30% glycerol (in HEMG buffer) gradient prepared in a 14 x 89 mm polyallomer centrifuge tube (Beckman Coulter). Tubes were centrifuged in an SW40 rotor at 4°C for 16 hr at 40,000 rpm. Fractions (0.550 ml) were collected and used in immunoblot analyses.

### Immunoblot

Protein was loaded onto Bis-Tris 4–12% gradient Novex gels and run for 150 V for 90 minutes. A wet transfer was performed for 2.5 hours at 165 mA at 4 degrees Celsius onto PVDF membranes. After transfer, membranes were blocked in 5% milk for 1 hour at room temperature before applying primary antibody (see below) and fluorescently labeled secondary antibodies for visualization using the LI-COR Odyssey.

### Antibodies Used

Subunit	Company	Catalog #
SMARCA4	Santa Cruz	sc-17796
ARID1A	Bethyl	A301-041
SMARCE1	Bethyl	A300-810A
SMARCB1	Santa Cruz	sc-166165
SS18	Cell Signaling	#21792
DPF2	Abcam	ab134942
SMARCD2	Santa Cruz	sc-101162
SMARCC1	Santa Cruz	sc-10756

(Continued on next page)

**Continued**

Subunit	Company	Catalog #
PBRM1	Bethyl	A301-591A
ARID2	Santa Cruz	sc-166117
BRD7	Santa Cruz	sc-376180
SMARCD1	Santa Cruz	sc-135843
BRD9	Abcam	ab66443

**Large-Scale Purification of Factors for Mass Spectrometry (C14orf80, C16orf59, C19orf25, TUBE1)**

After lentiviral introduction and selection, 293T cells with V5 tagged bait proteins were expanded to 15–20 confluent 15 cm dishes, or about 1E9 cells per preparation. Cells were scraped off the dishes and pelleted at 3000 rpm for 5 minutes at 4 degrees C. Cells were then resuspended in complete hypotonic buffer (10 mM Tris pH 7.5, 10 mM KCl, 1.5 mM MgCl<sub>2</sub>, 1 mM DTT, 1 mM PMSF, 1x protease inhibitor cocktail) and lysed for 5 minutes on ice. Lysate was then pelleted at 3000 rpm for 5 minutes at 4 degrees C and the upper cytoplasmic layer was collected, to which 3M KCl was added to a final concentration of 150 mM KCl, and rotated at 4 degrees C for 1 hour. After rotation, the cytoplasmic extract was spun at 20,000 RPM for 1 hour at 4°C in SW28 tubes in an ultracentrifuge. The lipid phase was removed and discarded, and the remaining extract was filtered through a 0.45 µm filter (Steriflip). This was then incubated with V5 beads and rotated overnight at 4 degrees C. Beads were then pelleted and washed 6 times with 12 mL of high salt buffer (50 mM Tris pH 7.5, 1 mM EDTA, 300 mM KCl, 1 mM MgCl<sub>2</sub>, 1% NP-40, 1 mM DTT, 1x protease inhibitor cocktail, 1 mM PMSF). After the last wash, beads were transferred into a 500 µL tube. For V5 elution, immunoprecipitations with V5 antibodies were eluted off of beads with 100 mM glycine pH 2.5 and quenched with 1M Tris pH 8.0 buffer (final 200 mM).

**Immunofluorescence (IF)**

Cells with overexpression of a tagged bait were split to 30–60% confluency onto a 24-well plate with appropriately sized and sterilized coverslips. Cells were allowed to adhere overnight. When cells were at the appropriate density and treatment time, cells were washed once with PBS in the plate and covered with either -20 degree Celsius methanol for 5 minutes (for C14orf80 and C16orf59 overexpression) or with 4% paraformaldehyde for 20 minutes (for C19orf25 overexpression), and then washed twice with IF wash buffer (0.1% NP40, 1 mM sodium azide, PBS 1X) and blocked overnight in blocking buffer (IF wash buffer + 10% FBS, filtered through 0.2µm filter). After blocking was complete, antibodies were diluted into blocking buffer and this solution was placed on the coverslips for 3 hours at room temperature. The dilutions were performed as follows:

Antibody	Manuf.	Cat.	Species	Dilution
Pericentrin	Abcam	ab4448	Rabbit	1:4000
V5	Thermo	R960-25	Mouse	1:1500
V5	Cell Signalling	#13202	Rabbit	1:3000
Anti-KDEL	Abcam	ab50601	Rat	1:300

This was followed by 3 washes with IF buffer (rinse, 5 minute incubation, repeat). Secondary antibody was diluted in blocking buffer 1:1000 and the slides were incubated for 1 hour. This was followed by a second round of 3 IF wash buffer washes (rinse, 5 minute incubation, repeat). Slides were then removed from the 24 well plate and mounted onto coverslips with mounting media containing DAPI stain (Prolong Gold [Invitrogen]).

Antibody	Manuf.	Cat.	Dilution
Goat anti-Mouse IgG Alexa Fluor Plus 555	Thermo	A32727	1:1000
Goat anti-Rat IgG Alexa Fluor 647	Thermo	A-21247	1:1000
Goat anti-Rabbit IgG Alexa Fluor 546	Cell Signaling	A-11010	1:1000

**Imaging and Analysis**

Images were captured at 60x magnification on a spinning disc confocal microscope. Images were taken as an 11 layer z-stack, which was then z-projected using the maximal value per pixel across stacks. ImageJ software was used for image processing and figure generation.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Using Fitness Thresholds to Define Genes for Analysis

Fitness screening data from RNAi screens in 501 cell lines (Tsherniak et al., 2017) and from CRISPR screens in 342 cell lines (Meyers et al., 2017) were downloaded from the Project Achilles Data Portal (<https://portals.broadinstitute.org/achilles>). Copy-number corrected versions of two additional CRISPR-Cas9 screening datasets, Wang et al. (Wang et al., 2017b) and GeCKO (Aguirre et al., 2016), were also used as comparisons (Meyers et al., 2017). Genes for downstream analysis were filtered for the presence of fitness effects upon genetic depletion across cancer cell lines: for both of the main datasets used in the paper (Project Achilles RNAi and CRISPR), genes were only included if the cell line most dependent on that gene exceeded a cutoff ( $-2$  for the RNAi dataset,  $-0.3$  for CRISPR dataset) and expresses that gene above a threshold of 0 RPKM. This resulted in a final set of 6300 genes in RNAi and 8161 genes in CRISPR. For parallel analyses involving gene expression (COXPRESdb Hsa3.c1-0, <http://coexpresdb.jp/download.shtml>), a union of the RNAi and CRISPR fitness genes was used.

### Fitness Profile Plots

For each of the selected genes in Figures 1B and 3B, their CRISPR-Cas9 fitness profiles were scaled between 0 and 1, where 0 represents the minimum essentiality and 1 represents the maximum essentiality of that gene across cell lines. A soft threshold was applied (scores were taken to the 3<sup>rd</sup> power), and both genes and cell lines were hierarchically clustered using complete linkage and correlation as a distance measure. Profiles were plotted using the ggjoy R package.

### Fitness Variation

For each gene in the RNAi and CRISPR datasets, the variance of that gene's fitness scores was calculated across all cell lines in which the gene was screened. For Figure S1B, genes were then binned by their membership or absence in the CORUM dataset (RNAi: No interaction,  $n = 4,850$ . Interaction: 1,455. CRISPR: No interaction,  $n = 6,502$ . Interaction,  $n = 1,659$ ). Significance was assessed using the Wilcoxon rank sum test.

### Enrichment of Correlated Gene Pairs among Protein-Protein Interactions

Pearson correlations between the fitness profiles of all gene pairs were calculated in both RNAi and CRISPR datasets, using pairwise complete observations only (i.e. only in cell lines where both genes were screened). For Figure S1C, genes present in the Core CORUM dataset of literature curated protein complexes were paired and broken into two bins: gene pairs from the same complex, and gene pairs from different complexes. The resulting distributions were compared using a two-sample KS test. In Figure S1D, for each protein-protein interaction dataset, each possible gene pair with was binned into one of two groups: interacting protein products or non-interacting protein products, and boxplots were then shown for these two groups across each of the available protein-protein interaction datasets. For Figure S1E, enrichment of protein-interactions among correlated gene pairs of varying strengths was calculated in the same way as previously done for pathways (Hart et al., 2017a). Correlated protein pairs were binned into 1000 bins based on their ranked correlation. The cumulative log likelihood of protein pairs to interact was calculated for all bins for all protein-protein interaction datasets. Log likelihoods for the first 100,000 ranked correlations are shown.

### Similarity Network Significance for Protein Complexes

For each correlation dataset (RNAi, CRISPR, COXPRESdb, GeCKO and Wang et al. and Costanzo et al. <http://thecellmap.org/costanzo2016/>) we performed a row-wise rank-transformation on the gene-gene correlation matrix, and symmetrized by taking the maximum rank between gene pairs.

For a given collection of genesets (Core CORUM, hu.MAP complexes, yeast complexes) and a similarity network at a given rank threshold, the significance of the observed internal-to-external edge density of each of the genesets in the collection is determined by calculating an empirical p-value using degree-preserved randomized networks. For the observed network and each of 10,000 randomized networks with preserved degree sequence, we calculated the ratio of the internal edge density to the external edge density (Figure S2) of each of the subgraphs of the genesets under consideration. We determined the empirical p-values representing the significance of each geneset as the fraction of shuffled networks for which this ratio in the null models exceeds that of the ratio in the observed network. We then apply a false discovery rate (FDR) correction to the p-values per correlation dataset. We used a FDR cutoff of 0.05 for significance.

To generate protein complex dataset recall plots, the cumulative percentage of protein complexes with significant fitness networks was plotted as a function of the rank threshold. We also plot the fraction of complexes recalled as a function of FDR for each correlation dataset.

To further characterize protein complexes that scored as significant in RNAi only ( $n = 29$ ), CRISPR only ( $n = 266$ ), or both ( $n = 199$ ) in Figure 2E, we assembled a set of gene features. Core essential genes were defined as the union of the CEG1 (Hart et al., 2015) and CEG2 (Hart et al., 2017b) datasets. Gene expression data was taken from the CCLE RNA-seq data for these cell lines (Barretina et al., 2012), and human-mouse dN/dS was obtained from BioMart (Smedley et al., 2015). Differences between the three groups of protein complexes were assessed with the Wilcoxon rank sum test.

For protein complexes that showed significant correlations in both RNAi and CRISPR datasets, we calculated the edge density ratio for that complex over all ranks in both datasets. Each protein complex was then assigned to either RNAi or CRISPR datasets as shown in Figure S3D, depending on which showed the largest edge density ratio for that complex over all rank thresholds.

## Network Visualizations

### Global Fitness Network Plots

Functional similarity networks were plotted in Cytoscape. In order to systematically choose rank thresholds for network visualizations for protein complex genesets in Figures S3E and 5C, we took the difference between the cumulative sum of internal edges (weighted by the inverse rank) and the cumulative sum of weighted external edges at each possible rank threshold, and chose the threshold that maximizes this difference.

### Individual Complex Similarity Plots

For individual complexes, network visualizations were generated using edges weighted by the inverse rank of the correlation. Functional similarity networks shown for individual complexes were shown with the top ranked correlation thresholds. For similarity networks highlighting shared subunits, we chose more lenient rank thresholds within the top 10% of ranked correlations. For coexpression networks, we chose a uniform rank 50 threshold, as many of the coexpression networks did not have internal edges to be shown at more stringent rank thresholds.

Networks were plotted with ggraph and ggiraph using the Fruchterman and Reingold force-directed layout algorithm as implemented in igraph R package. Node colors were chosen based on literature curated annotations (protein complex composition, functional modules, assembly components).

## Interface Size Dataset

Using the entire set of heteromeric protein complex structures in the PDB as of 2017-03-13, we identified polypeptide chains with >90% sequence identity to a human protein-coding gene. The sizes of all interfaces formed between pairs of subunits were calculated between all pairs of subunits using AREA-MOL as implemented in the CCP4 suite (Winn et al., 2011). For each pair of human protein-coding genes, the largest physical interface identified in the PDB was used for our analyses, and interactions with buried surface areas < 200 Å<sup>2</sup> were filtered from the dataset.

Structural heterodimers were defined to be protein pairs with at least 40% of each subunit's total surface area involved in the buried surface interface between subunits. A KS test was performed on the scaled rank correlations for each of the 271 structural heterodimers across the three datasets compared (RNAi, CRISPR, COXPRESdb).

## Modularity Overlap Enrichment

Protein complexes that had significant fitness networks in either RNAi or CRISPR and between 10 to 150 subunits were considered for modularity analysis. For each of these 133 protein complexes, we generated networks using the following structural edgelist (Bioplex 2.0, PDB, and huMAP), as well as genetic edgelist (RNAi or CRISPR, CoexpresDB) with optimal thresholds chosen as stated above.

For each protein complex, we performed Louvain community detection as implemented in the igraph R package across networks derived from these datasets. We then assessed overlap significance between module assignments using two-sided Fisher's exact tests. For each of the three datasets shown (Coexpression, Fitness, and huMAP), networks were compared to available PDB or Bioplex physical networks. FDR correction was applied across p-values in each of the three groupings, and a threshold of FDR < 0.2 was used.

## Heatmaps and Clustering

Hierarchical clustering was performed on fitness profile correlation matrices using the Euclidian distance metric and complete linkage clustering. Heatmaps were visualized using the heatmap.2 function in the gplots R package.

## Scoring Significant hu.MAP Complexes

Predicted protein complexes from hu.MAP were obtained from their website (<http://proteincomplexes.org/download>). The hu.MAP interaction network is weighted by a predicted pairwise probability of interaction between each protein, ranging from 0.75 (the significance cutoff) to 1 (maximum probability). In order to score protein complexes by a combination of physical interaction evidence and correlation strength, we took the pairwise product of the CRISPR correlations and predicted interaction probability for all gene pairs, and ranked significant complexes by the average of those products. We then selected protein complexes for validation that included one or more subunits of uncharacterized function. The top 300 protein complexes by this scoring paradigm were included as a supplement (Table S6).

## Gene Evolution Tables

The following ENSEMBL gene trees were used as references for Figures 5I and S9H:

**C14orf80:** ENSGT00390000011474

**C16orf80:** ENSGT00390000011149

**RINT1:** ENSGT00390000017006

**NBAS:** ENSGT00390000012474  
**ZW10:** ENSGT00390000016427  
**C19orf25:** ENSGT00390000007991  
**ZWILCH:** ENSGT00390000013696  
**ZWINT:** ENSGT00390000017639  
**KNTC1:** ENSGT00390000007883

#### DATA AND SOFTWARE AVAILABILITY

All code used to generate figures in the manuscript are available through an online repository: <https://github.com/robinmeyers/pan-meyers-et-al>. All data from this manuscript are available at: <https://doi.org/10.6084/m9.figshare.6005297>.